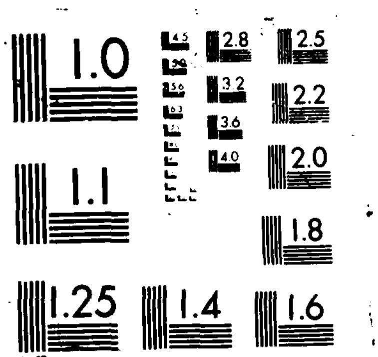
AD-A187 294 ENVIRONMENTAL GRADIENT ANALYSIS ORDINATION AND CLASSIFICATION IN ENVIRONM..(U) CONSTRUCTION ENGINEERING RESEARCH LAB (ARMY) CHAMPAIGN IL 1/2

UNCLASSIFIED A J KRZYSIK SEP 87 CERL-TR-N-87/19 F/G 6/6 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

USA-CERL TECHNICAL REPORT N-87/19
September 1987
Training Area Impact Prediction

US Army Corps of Engineers
Construction Engineering Research Laboratory

AD-A187 294

Environmental Gradient Analysis, Ordination, and Classification in Environmental Impact Assessments

by
Anthony J. Krzysik

This report develops the theoretical foundation for analytical description and quantification of habitat structure. The methodology is based on the concept of environmental gradient analysis. The analytical description of environmental gradients is shown to be an eigenanalysis problem, mathematically equivalent to the largest eigenvector (or first principal component) of a principal components analysis. The analytical representation of an environmental gradient, itself a *single* variable, is empirically demonstrated to have similar ecological information as the combination of all the original 58 habitat variables describing five Mojave Desert study sites.

Two vastly different data bases were analyzed to explore the effects of sample sizes and variable selection on the ordination of study sites in both principal components and canonical variate space. Merits and shortcomings of principal components analysis, canonical analysis of discriminance, and cluster analysis for the ordination and classification of samples are reviewed in detail. An optimal "recipe" is developed for ordinating habitat samples along the major environmental (disturbance) gradient.

Canonical variate ordinations were shown to be sensitive to sample sizes, the choice of variables, and linear dependency among variables. Canonical analysis of discriminance is a very effective mechanism for classifying samples into *a priori* established groups, or for identifying variables that contribute significantly to group discrimination.

Cluster analysis proved to be an excellent technique for exploring the effectiveness of variable subsets in classifying samples into *a priori* established groups. The R-mode of cluster analysis is recommended for identifying independent variable subsets and eliminating redundant variables.

USER EVALUATION OF REPORT

REFERENCE: USA-CERL TR N-87/19, *Environmental Gradient Analysis, Ordination, and Classification in Environmental Impact Assessments*

Please take a few minutes to answer the questions below, tear out this sheet, and return it to USA-CERL. As a user of this report, your customer comments will provide USA-CERL with information essential for improving future reports.

1. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)

2. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.)

3. Has the information in this report led to any quantitative savings as far as man hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.

4. What is your evaluation of this report in the following areas?

   a. Presentation:

   b. Completeness:

   c. Easy to Understand:

   d. Easy to Implement:

   e. Adequate Reference Material:

   f. Relates to Area of Interest:

   g. Did the report meet your expectations?

   h. Does the report raise unanswered questions?

i. General Comments. (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.) ____________

____

____

____

____

5. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

Name: ____________

Telephone Number: ____________

Organization Address: ____________

____

____

6. Please mail the complete form to:

Department of the Army
CONSTRUCTION ENGINEERING RESEARCH LABORATORY
ATTN: CERL-IMT
P.O. Box 4005
Champaign, IL 61820

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A187294

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704 0188 Exp Date Jun 30 1986

1a REPORT SECURITY CLASSIFICATION: Unclassified

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S): USA-CERL TR N-87/19

5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION: U.S. Army Construction Engr Research Laboratory

6b OFFICE SYMBOL (If applicable): CECER

7a NAME OF MONITORING ORGANIZATION

6c ADDRESS (City, State, and ZIP Code): P.O. Box 4005, Champaign, IL 61820-1305

7b ADDRESS (City, State, and ZIP Code)

8a NAME OF FUNDING / SPONSORING ORGANIZATION: OCE

8b OFFICE SYMBOL (If applicable): DAEN-ZCF-B

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c ADDRESS (City, State, and ZIP Code): 20 Massachusetts Ave., N.W., Washington, DC 20314-1000

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
4A76270	A896	A	026

11 TITLE (Include Security Classification): Environmental Gradient Analysis, Ordination, and Classification in Environmental Impact Assessments (Unclassified)

12 PERSONAL AUTHOR(S): Krzysik, Anthony J.

13a TYPE OF REPORT: Final

13b TIME COVERED: FROM ________ TO ________

14 DATE OF REPORT (Year, Month, Day): 1987, September

15 PAGE COUNT: 122

16 SUPPLEMENTARY NOTATION: Copies are available from National Technical Information Service, Springfield, VA 22161

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
06	06	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

environmental gradient analysis
environmental impact assessment
ecology
statistical analysis

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This report develops the theoretical foundation for analytical description and quantification of habitat structure. The methodology is based on the concept of environmental gradient analysis. The analytical description of environmental gradients is shown to be an eigenanalysis problem, mathematically equivalent to the largest eigenvector (or first principal component) of a principal components analysis. The analytical representation of an environmental gradient, itself a *single* variable, is empirically demonstrated to have similar ecological information as the combination of all the original 58 habitat variables describing five Mojave Desert study sites. A common example of an environmental gradient is a series of study sites ranked on the extent of habitat disturbance or on successional development.

(cont'd)

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: D. P. Mann

22b TELEPHONE (Include Area Code): (217) 373-7223

22c OFFICE SYMBOL: CECER-IMT

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

Block 19, cont'd:

Two vastly different data bases were analyzed to explore the effects of sample sizes and variable selection on the ordination of study sites in both principal components and canonical variate space. Merits and shortcomings of principal components analysis, canonical analysis of discriminance, and cluster analysis for the ordination and classification of samples are reviewed in detail, including raw data refinement (transformations, standardization), the structure of input matrices, and orthogonally rotated solutions. An optimal "recipe" is developed for ordinating habitat samples along the major environmental (disturbance) gradient. This gradient is defined by the first principal component (PC 1). The principal components analysis uses a Pearson correlation matrix of appropriately transformed habitat variables for input, with subsequent varimax rotation transforming the n x 4 factor pattern matrix. Additionally, this specific solution enhances the ecological interpretation of PC 2, PC 3, and PC 4. The concept of environmental gradient analysis and the specific methodology that is proposed is not only robust and possesses extremely broad applicability and flexibility for environmental assessments in any kind of ecosystem, but can also be used to accurately quantify species-habitat associations. The proposed methodology can also be applied to land-use management, maintenance and monitoring programs, and to environmental impact prediction and mitigation.

Canonical variate ordinations were shown to be sensitive to sample sizes, the choice of variables, and linear dependency among variables. Canonical analysis of discriminance is a very effective mechanism for classifying samples into *a priori* established groups, or for identifying variables that contribute significantly to group discrimination.

Cluster analysis proved to be an excellent technique for exploring the effectiveness of variable subsets in classifying samples into *a priori* established groups. The R-mode of cluster analysis is recommended for identifying independent variable subsets and eliminating redundant variables.

FOREWORD

This research was performed for the Assistant Chief of Engineers, Office of the Chief of Engineers (OCE), under Project 4A76270A896, "Environmental Quality of Military Facilities"; Technical Area A, "Installation Environmental Management Strategy"; Work Unit 026, "Training Area Impact Prediction." The work was performed by the Environmental Division (EN) of the U.S. Army Construction Engineering Research Laboratory (USA-CERL). Mr. Donald Bandel, DAEN-ZCF-B, was the OCE Technical Monitor.

Mr. David Tazik and Mr. Robert Szafoni assisted with the collection of the Illinois habitat data. Field assistance was provided by Mr. Charles Facemire, Ms. Barbara Zupcic, Ms. Diane Szafoni, and Mr. Steven Railsback in Illinois, and Mr. Joseph Burke and Mr. Thomas Raney in the Mojave Desert. The assistance of the following people is also acknowledged: Mr. Robert Riggins, Dr. Edward Novak, Dr. R. K. Jain (all of USA-CERL); Dr. William Lower and the field crew of the Environmental Trace Substance Research Center (University of Missouri); Mr. Alan Waite, Mr. John Carroz, MAJ Robert Schwegler, MAJ Donald Dickenson, LT Daniel Danarski (all of Fort Irwin); Mr. Benjamin Gaudian, Mr. Charles Goodsen, and Mr. Lewis Butcher (all of Goldstone); and Dr. Edwin Herricks, Dr. James Karr, Dr. Lowell Getz (all of the University of Illinois).

Dr. R. K. Jain is Chief of USA-CERL-EN. COL Norman C. Hintz is Commander and Director of USA-CERL, and Dr. L. R. Shaffer is Technical Director.

CONTENTS

CONTENTS (Cont'd)

FIGURES

FIGURES (Cont'd)

FIGURES (Cont'd)

TABLES

TABLES (Cont'd)

ENVIRONMENTAL GRADIENT ANALYSIS, ORDINATION, AND CLASSIFICATION IN ENVIRONMENTAL IMPACT ASSESSMENTS

1 INTRODUCTION

Background

Traditionally, the physical structure (physiognomy) of vegetation is the most important habitat describer, especially for understanding the structure and species richness of bird communities. Previous studies (70, 80, 52, 60, 91, 2, 106, 119, 111, 112, 86, 58, 73, 16)* have emphasized the importance of foliage height diversity, vertical layers of vegetation, horizontal patchiness (habitat mosaics, including the presence of habitat edges), and optimal "protected" nest sites (e.g., spinescent shrubs; tree, snag, and cactus boles).

Plant species compositions (floristics) have also explained the makeup of avian communities (46, 6, 118, 69, 32, 50, 54, 114). Since a plant species may have a specific physical structure, it is generally difficult to uncouple species composition and physical structural components. However, specific plant species often show sufficiently distinct resources (e.g., fruit, nuts, seeds, nectar, pollen, specific nest sites) that the value of these resources can be quantified. Avian researchers have generally overlooked plant species, in terms of the arthropod communities they harbor, as specific food resources for birds. Arthropod species richness, taxonomic compositions, numerical abundances, and body size distributions may depend more on the taxonomic composition of the vegetation than on vegetation density, volume, or structural complexity. Furthermore, the evolution of plant toxins and the corresponding arthropod herbivores is closely interrelated (10, 24, 97, 27, 28, 29, 68, 81, 38, 85, 8). However, Jermy (57) presents evidence that strongly challenges traditional theories of the coevolution of plant defense toxins and insect herbivory. Plants' chemical defenses influence arthropod communities and therefore avian communities in two important ways. First, plant species that have toxins to deter herbivores may harbor fewer numbers and/or smaller body sizes of arthropods. Theoretically, these plants should maintain fewer taxa of herbivores, since specialization is required to detoxify chemical defenses. Second, at least some of the arthropod fauna may be unpalatable to avian predators.

Both leaf morphology and the arrangement of leaves on branches correlate with avian foraging strategies, and therefore represent species-specific resources (51). Bark surface morphology (55, 56, 107) and twig diameter (77) also determine species-specific foraging on bole and limb surfaces.

These examples show that many environmental features--all potentially important and some very critical--underlie habitat characterization. Describing habitats in an analytical framework, as well as specifically identifying the relative importance of habitat components to the community structure of

*Numbers in parentheses refer to references provided on pp 107-114.

vertebrate faunas, challenges experimental designs, sampling theory, data analysis, and ecological interpretation. An even stronger challenge faces the synthesis of ecological theory and land, resource, and wildlife management.

Research that focuses on understanding the complexity of interrelationships between species and their habitat requirements is necessary for maintaining, restoring, and managing Army lands to optimize training realism, conserve natural resources, maintain ecosystem integrity, and comply with Federal and State mandated environmental regulations. Previous research by the U.S. Army Construction Engineering Research Laboratory (USA-CERL) at the Army National Training Center, Fort Irwin, CA, collected data on the effects of large-scale Army training maneuvers and war games on the surrounding desert ecosystem (64). This study also developed analytical methodologies for quantifying environmental impact assessments, and analytically described and summarized species-habitat associations. The information obtained will be used as guidelines for managing desert nongame species and ecological communities at Army installations.

An important basis for the previous USA-CERL study (64) was the concept of environmental gradients, which ranks impacts to provide a base for determining land and species disturbances. The information in the current report outlines the theoretical and analytical framework used for the previous study.

Objective

The objectives of this report were to: (1) present the concept of environmental gradients and evaluate its potential for environmental impact assessment, (2) develop an analytical methodology for describing environmental gradients, (3) explore the sensitivity of sample sizes and variable selection on ecological ordinations, and (4) explore analytical methodologies for classifying environmental/ecological samples.

Scope

A consequence of this research is the analytical and unbiased description of habitat structure in terms of physical and vegetation characteristics. In some instances, chemical parameters may be important: pH, salinity, alkalinity, oxygen tension, micro- and macronutrients, allelopathics, pesticides, toxins, and heavy metals compositions and concentrations. However, even in these circumstances, a knowledge of vegetation and soils will usually be closely correlated with the chemical environment in terrestrial as well as riparian ecosystems. The chemical environment plays a more immediate and direct role in the spatial and temporal organization of aquatic ecosystems. Chronic dose concentrations of natural or anthropic elements or compounds will not be considered as part of the habitat *Gestalt*.

Approach

Two databases (Mojave Desert and Illinois) were selected to explore the effects of sample sizes and variable selection on study site ordinations for gradient characterization. Variables for quantifying habitat characteristics

were chosen for the two database locations. Data were gathered for each of the variables and subjected to multivariate statistical analysis. Analytical results were then examined and the advantages and disadvantages of each stastistical analysis method used were noted.

Mode of Technology Transfer

This research contributes to the fundamental understanding of Army training impacts and their quantitative assessment and description. It is recommended that the results be used to develop more effective impact prediction methods, land maintenance technologies, and analytical approaches to resource and wildlife management.

2 THEORETICAL BASIS FOR CHARACTERIZING ENVIRONMENTAL GRADIENTS

Environmental Gradients

In any kind of environmental assessment or evaluation, a series of "study sites" is selected that represents an ordinal ranking of impacts, disturbances, or recovery, including controls. This means that, based on visual perception and experience, an ecologist, land manager, or concerned citizen identifies land values on an environmental ranking (an environmental gradient). Typically, the two most familiar examples of these gradients are environmental disturbance and successional gradients. After sites are identified, experience and the literature dictate the choice of sampling strategies and which habitat (environmental) variables to measure. Given that information, samples of this specific nature must exemplify their greatest intersite variance (variability) on habitat-variable sets that closely associate with habitat description. What is the optimal set of habitat variables and the relative importance of each variable that best quantifies or describes habitat structure in the context of ordinal environmental gradients? This report derives the mathematical form of environmental gradients and shows it to be a generalized eigenanalysis solution of the well known principal components* solution.

Canonical Analysis of Discriminance

Terminology in discriminant analyses has been used inconsistently and inappropriately. According to Pimentel (78), the analysis used here is canonical analysis of discriminance (CAD). This analysis weights predictor (habitat) variables such that their linear combinations maximally distinguish (discriminate) among two or more predetermined groups or classes. The F-ratio tests the criterion for measuring class differences (65, 79). Basically, $F = SS_b/SS_w$. By rewriting sums of squares terms in the form of vectors of linear combinations of predictor variables, the matrix form reduces to:

$$\mathbf{v}'\mathbf{B}\mathbf{v}/\mathbf{v}'\mathbf{W}\mathbf{v} = \lambda \quad \text{(see Eq 37)}$$

Fisher (31) was the first to propose λ as the discriminating criterion. The discriminant problem reduces to extracting the set of weights that maximizes λ. This is accomplished by the Lagrange multiplier (see Eq 38). For more details, refer to Tatsuoka (102), Pimentel (78), or Dillon and Goldstein (17). Tatsuoka (101) presents an unusually elementary and lucid introduction to discriminant analysis. Anderson (4) and Seber (90) provide advanced treatments.

Discriminant analyses appear to be potentially valuable tools, since the releative importance or lack of significance of each habitat variable can be evaluated according to its ability to separate treatments (impacts) from controls. However, discriminant analyses pose problems.

---

*A dimension reduction ordination technique.

Discriminant analyses have many practical undesirable features, despite their theoretical advantages and widespread use by researchers. Several publications (66, 108, 62, 110, 115, 116, 4, 90) discuss the shortcomings of discriminant analyses concerning statistical assumptions, sampling variability, and analysis interpretation. Although they are frequently used in ecological studies, stepwise discriminant procedures have been criticized, since many environmental variables are ususally highly correlated (25, 40). Seber (90) presents a good review of variable selection in discriminant analysis and emphasizes that stepwise procedures strongly depend on the assumptions of normality and homogeneous dispersion matrices. The most commonly used statistical packages employing discriminant procedures for selecting variable subsets also depend on these assumptions (42). Not only are these assumptions difficult to test, but they may not even be tenable. Furthermore, ecological data, even transformed, rarely meet the above assumptions. Hawkins (47) has developed a procedure to test simultaneously for homogeneity in the covariance matrices and multivariate normality, based on the Anderson-Darling test statistics (22). However, my experience has been that stepwise discriminant analyses are totally inappropriate for the selection of relevant and interpretable variables, unless the underlying ecological pattern is very simple and a relatively small number of independent (predictor) variables are involved.

The procedure for selecting variable subsets in this research was a two-level process. First, variables were identified by their ability to quantify the major environmental gradient present in the study sites. The relative quantitative value of predictor variables was determined by *a priori* orthogonal contrasts, an ANOVA routine (64). Second, various logical subsets of these variables (based on experience) were subjected to CAD. Inspection of these analyses results clearly identified the subset of variables that not only defined a prominent environmental impact gradient on the first canonical variate axis and produced a good separation of study sites in canonical space, but was ecologically interpretable on the basis of first-hand field experience.

The number of discriminating variables is limited to the number of sampling units. Generally, this is not a problem, unless the researcher samples only several transects or quadrants.

Cluster Analysis

Cluster analysis is a powerful analytical tool for organizing samples, observations, or objects into groups based on predictor (independent) variables (1, 94, 45, 26, 21, 67, 84, 90). Clustering procedures are generally hierarchical. In other words, the resulting classification produces ranked classes where the members of inferior ranking clusters become members of larger higher-ranking clusters. The ranking produces the well known dendrogram tree. This is a logical procedure for taxonomic data and for visualizing the underlying pattern of ranked relationships. Nonhierarchical methods are potentially useful in ecology (67), but are more difficult to compute and not readily available in data analysis software packages. These methods optimize within-group homogeneity, rather than hierarchical routes between objects and major clusters. Objects are clustered with groups only on the basis of their direct relationships, without regard to hierarchical affinities.

There are two approaches for hierarchical classification of n objects. Using divisive methods, n objects begin in a single group and are separated into two groups, which in turn are separated into four, etc., until only a pair of objects remain. Divisive methods may be monothetic or polythetic. Monothetic algorithms consider a single descriptor variable at a time. This may lead to erroneous classifications (39a, 15). Polythetic algorithms use an association matrix which considers the complete combination of many variables. However, for realistic data (even when the number of objects and predictor variables is moderate) polythetic methods require unrealistic computer effort (67). Seber (90) discusses advantages and disadvantages of divisive techniques. The other approach (agglomerative) begins by examining all individual objects and pairs the two most similar, then the next pair that is most similar, merging groups accordingly until a single group comprising n objects is formed. Agglomerative methods are much more widely used and highly recommended (94, 67, 84). Table 7.9 of Legendre and Legendre (67) provides a useful synoptic summary of clustering methods.

There are six commonly used hierarchical agglomerative clustering algorithms for mainframe computers: (1) the unweighted pair-group method that uses arithmetic averages (UPGMA), (2) the weighted pair-group method that uses arithmetic averages (WPGMA), (3) the single-linkage clustering, (4) the complete linkage clustering, (5) the centroid method, and (6) Ward's method (NTSYS,* CLUSTAN [120], SAS [87], BMDP [18], SPSS[x] [98]). The first five of these algorithms actually fit a single general model, and therefore, a single computer program can be used where four parameters are varied to select the desired algorithm (66a, 66b). Algorithms are executed on a similarity or resemblance matrix. This matrix consists of coefficients (similarity, dissimilarity, distance, etc.) that reflect the degree of association (or dissimilarity) among all possible pairs of the samples, observations, or objects to be classified. There is a very wide variety of resemblance coefficients in use, including various correlation coefficients, cosines of variable vectors, and distance measures. Common distance measures include Euclidean distance (usually its square), Manhattan metric, and arccosine of the correlation. The literature provides discussions and suggestions for selecting and using resemblance coefficients (92, 14, 1, 94, 83, 15, 19, 45, 7, 9, 26, 21, 67, 84). The variables or attributes describing the samples may be measured on nominal, ordinal, ratio, or interval scales. Counts, frequencies, or even presence/absence binary data are all equally appropriate. At least 12 different qualitative resemblance coefficients have commonly been used for presence/absence data. Cluster analyses are routinely conducted that mix several variable scales. It is common practice in cluster analysis to transform and/or standardize raw data values.

Cluster analysis is a very useful classification technique for grouping observations, samples, species, study sites, objects, etc. (e.g., dependent or criteria variables) on the basis of independent or predictor variables (e.g., environmental, morphological, etc. variables). This is the most frequent use of cluster analysis, and is termed Q-analysis. The independent or "description" variables can be clustered similarly on the basis of

---

*Dr. F. James Rohlf, Department of Ecology and Evolution, The State University of New York, Stony Brook, New York, 11794.

dependent or sample variables. This is known as R-analysis. R-analysis is useful for reducing the dimensionality of experimental data sets.

The major advantages of cluster analysis are its inherent simplicity and its very broad applicability to any kind of data, sample sizes, etc. However, there are drawbacks. The researcher must choose from among a variety of clustering algorithms, and each method usually yields different (sometimes very different) results. There is an even larger choice when constructing a resemblance matrix.* Again, the results are sensitive to the similarity or distance coefficients employed. There will probably never be a "correct" or optimal methodology for cluster analysis (117, 76). Most statisticians agree that in cluster analysis statistical inference or hypothesis testing may be tenuous. However, Dillon and Goldstein (17) and Romesburg (84) do not agree with this consensus.

An excellent application of cluster analysis is exploring the effectiveness of independent variable subsets on classifying entities of known group affiliation, as done in this study. Here, the choice of a clustering algorithm and the form of the resemblance matrix was considered very carefully on the basis of previous experience and the literature (82, 93, 94, 67, 84, 90). (Chapter 4 provides details.)

Cluster analysis proved highly beneficial, and can be recommended for exploring the effectiveness of various independent variables to classify sampling units of known group affiliation.

Cluster analyses using the Mojave Desert data set (described in Chapter 4) were performed to contrast this technique with principal components and canonical variate ordinations, and to illustrate some features of eigenanalysis and discriminant analysis.

Cluster analysis was also used to empirically demonstrate that the use of a reduced n x 4 factor pattern matrix produces rotated principal components solutions that are ecologically more realistic and interpretable than solutions derived from the full n x n matrix.

*The resemblance matrix is developed from the raw data matrix and consists of similarity, dissimilarity, distance, or correlation coefficients, directional cosines or some quantifiable measure of the "resemblance" or association of data matrix elements. The resemblance matrix is the input for the clustering algorithm.

3 DATABASE SELECTION

Two vastly different databases were analyzed to explore the effects of sample sizes and variable selection on the ordination of study sites in both canonical variate and principal components space. The importance of this evaluation cannot be overemphasized. In the literature, researchers initiate important management decisions on the basis of inferences from their results without reference to the process of variable selection (e.g., they do not state how or which variables were selected for their analyses). Thus, it is not clear if their results, statistical inferences, or ecological interpretations were sensitive to variable selection.

For this study, the two databases used for analytical habitat description were: (1) 23 linear transects associated with Army training impacts at Fort Irwin, California, in the central Mojave Desert (35^{o} 38' - 35^{o} 8' N, 116^{o} 56' - 116^{o} 19' W) and (2) 369 square quadrats associated with strip-mine succession in east central Illinois.

The central Mojave Desert consists of broad valley plains between rugged mountain ranges and occasional high plateaus and dry lake beds (playas). Away from the mountains are alluvial fans or bajadas (coalesced fans) that gradually slope into the valleys. Alluvial substrates are gravelly or rocky since they are formed by the erosional breakdown of the mountains. Boulder fields or large rubble can be found at the base of the mountains. Washes occur throughout the valleys, bajadas, and alluvial fans.

USA-CERL Technical Report N-85/13 (64) provides map coordinates and elevations for all of the Mojave Desert study sites. Appendix A provides general ecological characteristics of the study sites.

The nine sites in Illinois, called B, C, D, E, F, G, H, I, J, were all within 5 km of one another, and were located in Kickapoo State Park and Leisure Time Estates in central Vermilion County, east central Illinois, 20 km west of the Indiana state line. Another site (site A), a University of Illinois natural area (Phillips Tract) was located in Champaign County, 35 km west of the other sites. Appendix D gives general ecological characteristics and sizes of the Illinois study sites. Most of the study sites were 10 ha.

Each transect or quadrat in the two areas will be referred to as a sampling unit. The Illinois data set is much larger, since it has 16 times more sampling units. However, the Mojave data set is of "higher statistical quality." This terminology is not meant to imply better or more accurate data, but is based on two factors: (1) sample sizes to obtain habitat parameters within each Mojave transect were usually larger than those used to obtain comparable parameters in the Illinois quadrats, and (2) the size of the sampling area used to measure habitat variables was 57 times larger on the transect sampling units (16 ha) than on the quadrat sampling units (0.28 ha). Therefore, by sampling more intensely and by covering a much larger area to obtain these samples, the estimated habitat parameters should more accurately represent the entire study site.

These two experimental designs address an interesting question. If two sampling designs have equal intensities of sampling, but one site contains

many small plots with few samples per plot and the other site uses larger plots with large numbers of samples per plot, will the multivariate representations in both designs be comparable? This report addresses this problem from the statistical viewpoint. However, this question may be more relevant to biological/ecological reality than to statistical theory. Several examples will clarify this point.

The predominant habitat component of the Mojave Desert is the three-dimensional distribution of creosote bush (Larrea tridentata). Burroweed (Ambrosia dumosa) density or cover and substrate particle size are other important components. Therefore, the habitat structure is relatively simple, and is very similar from place to place (low α diversity). Extreme aridity results in low spatial and temporal productivity, so vegetation is sparse and patchy on a scale of 10 to 50 m. Therefore, fewer but larger sampling units are necessary.

Contrastingly, regions with higher rainfall have not only more dense, but much more complex vegetation structure, with a variety of trees, shrubs, vines, and ground cover. A high degree of vegetation complexity increases the potential for high intrasite variability (high α diversity). For example, a large windblown tree opens the canopy in an eastern forest, and blackberry, raspberry, or rose thickets develop along with grape vines. Similarly, during a series of unusually wet years, local depressions in a forest may collect enough water to saturate the soil and kill flood-intolerant tree species, thus producing snags. Both of these patches in the forest represent ecological opportunities (food and nest sites) that may be unavailable or rare in the closed canopy forest interior. Therefore, when habitat complexity is greater, sampling must be done on a smaller scale to quantify habitat mosaics, particularly if an important feature of the research is to decipher species-habitat relationships. Furthermore, if primary productivity is high and the vegetation dense, smaller sample sizes may be sufficient for quantification. Therefore, assuming that time and manpower are limited, as they always are, specific experimental designs for habitat quantification are a function of habitat complexity or primary productivity.

4 DESCRIPTION OF SAMPLING VARIABLES AND DATA COLLECTION AND ANALYSIS METHODS

This chapter describes the variables chosen for quantifying habitat characteristics at each study site. Also described are the methods used to gather data for each variable and the data analysis methods.

Mojave Desert Study Sites and Habitat Variables

Five study sites were located at Fort Irwin. These sites were chosen to represent three levels or degrees of habitat disturbance (severely impacted, moderately impacted, lightly impacted) and two unimpacted control sites. Typically, vehicle traffic and Army training activities occur in the valleys, and impacts decline as one progresses into the alluvials toward the mountains. The severely impacted site (S) was in a broad valley about 600 m from the main road. The moderately impacted site (M) was on the alluvial about 3 km from the S site. Respective control sites (valley control [VC] and alluvial control [AC]) were located at Goldstone, an unimpacted portion of Fort Irwin, leased to the National Aeronautics and Space Administration/Jet Propulsion Lab for tracking deep space satellites. A fifth study site (L) was lightly impacted and located on a high desert plateau.

Three to six 800-m-long linear transects (sampling units) were located in each of the five study sites (Appendix A, N = 23*). Four additional transects for sampling habitat variables were located along each of the 800-m sampling units (N = 4 x 23 = 92). These habitat transects had their origins at the 100-, 300-, 500-, and 700-m loci of the 800-m transects. The compass bearing of the habitat transect was randomly determined by casting dice and employing a technique such that all integers from 0 through 359 had an equal probability of occurrence. Habitat transects were 100 x 4 m (50- x 4-m transects were used when shrub density was high). Each woody plant (height $\geq$ 0.2 m) whose center fell within the transect was identified, and its average diameter (measured at its maximum horizontal projection) and maximum height were recorded to the nearest 0.1 m.

Desert shrub individuals, particularly creosote bush, are not easily delimited (23, 61). Creosote bush scrub consists of circular or semicircular bands of clonal growth patterns. The center of the shrub dies, but is replaced on the periphery by new stem growth from the roots. The resulting growth pattern is a ring of circular or elliptical satellite clumps with a barren center (99, 109). Therefore, delineation of shrub individuals was somewhat subjective. Shrubs were judged as individuals when their foliage was separated by at least 10 cm or when their respective radii were clearly outlined (e.g., burroweed foliage occurs in dense, compact, spherical clusters).

Each vegetation transect was traversed, and every 2 m a sighting tube with crosshairs was used to record 50 "ground hits" from the following categories: grass, forb (nonwoody annual and perennial vegetation), litter

*Total of transects for all five sites was 23.

(dry and decaying plant material lying on the ground surface), sand (< 3 mm), coarse sand (3 mm to < 1 cm), gravel (1 to < 8 cm), and rock ($\geq$ 8 cm).

Shrub density was calculated as:

$$I_{xs} = (10^4/16J_x) \sum_{j=1}^{J_x} \sum_{k=1}^{4} N_{xsjk}/L_{xjk} \qquad \text{[Eq 1]}$$

where:

I_{xs} = density of species (or height category) s at site x (shrubs/ha)
J_x = number of sampling units at site x
N_{xsjk} = number of individual shrubs in the k^{th} habitat transect of the j^{th} sampling unit of species s at site x
L_{xjk} = length of the k^{th} habitat transect of the j^{th} sampling unit at site x.

Shrub cover was calculated as:

$$C_{xs} = (10^4/4) \sum_{i=1}^{N_{xs}} \pi(D_{xsi}/2)^2 / \sum_{j=1}^{J_x} \sum_{k=1}^{4} L_{xjk} \qquad \text{[Eq 2]}$$

where:

C_{xs} = cover of species (or height category) s at site x (m^2/ha)
D_{xsi} = diameter (m) of the i^{th} individual of species s at site x
N_{xs} = sample size of species s at site x.

Ground cover was calculated as:

$$G_{xg} = (1/2J_x) \sum_{j=1}^{J_x} \sum_{k=1}^{4} P_{xgjk} \qquad \text{[Eq 3]}$$

where:

G_{xg} = percent ground cover of g (grass, forbs, or litter) at site x
P_{xgjk} = number of "hits" of g in the k^{th} habitat transect of the j^{th} sampling unit at site x.

Substrate particle size distribution was calculated as:

$$S_{xp} = \sum_{j=1}^{J_x} \sum_{k=1}^{4} P_{xpjk} / \sum_{p=1}^{4} \sum_{j=1}^{J_x} \sum_{k=1}^{4} P_{xpjk} \qquad \text{[Eq 4]}$$

where:

S_{xp} = percent of surface substrate of particle size p at site x
P_{xpjk} = number of "hits" of particle size p in the k^{th} habitat transect of the j^{th} sampling unit at site x.

Horizontal heterogeneity was defined as the *mean within-transect heterogeneity* of three shrub parameters: diameter, height, and density. Within-transect heterogeneity was the coefficient of variation calculated from the four habitat transects. The mean from all 800-m-long sampling units at a given site therefore determined horizontal heterogeneity.

Habitat heterogeneity was calculated as:

$$H_{xp} = (1/J_x) \sum_{j=1}^{J_x} SD_{xpj}/\bar{X}_{xpj} \qquad \text{[Eq 5]}$$

where:

H_{xp} = index of horizontal heterogeneity of parameter p
p = shrub diameter, height, or density
SD_{xpj} = standard deviation of parameter p in the j^{th} sampling unit at site x
$\bar{X}_{xpj}$ = the mean of parameter p in the j^{th} sampling unit at site x.

Appendix B lists all Mojave Desert habitat variables. Appendix C defines the transformations for these variables.

Illinois Study Sites and Habitat Variables

Each study site (with the possible exceptions of A and F) was very carefully selected so that it occupied the center of a much larger area that had habitat identical to the respective study site. Thus, both island and edge effects were eliminated. All sites were gridded into 0.28-ha square plots (52.7 m per side). Each 0.28-ha plot was a sampling unit (Figure 1).* Although it was desirable to have square study sites, as Figure 1 suggests, this was not always feasible (e.g., bottomland forest sites). A circular 26-m-diameter quadrat (0.053 ha) was established at the center of each 0.28-ha plot to quantify habitat structure. The method for quickly delineating circular quadrats without using a tape measure or markers is given in James and Shugart (53). The following sampling scheme was used in each 26-m circular quadrat. At each of the 20 sampling points "A" and "B" (Figure 2), a heavy cardboard tube fitted with crosshairs was used to determine the presence/absence of vegetation for ground cover (vegetation < 1 m high) and canopy cover (vegetation > 2 m high). To eliminate focus bias, the ocular was positioned at hip level or above the head before looking through it. A hit was scored as positive if vegetation was observed at the intersection of the crosshairs. Also recorded at the same time with this method was the presence/absence of aquatic habitats and whether the ground was level or sloped. At each of the 12 sampling points "B," the presence/absence of vegetation at 11 vertical levels was determined, based on the intersection of vegetation with a 7.5-m telescopic measuring rod. Levels greater than 7.5 m were estimated. Appendix E gives the height categories used.

*Tables and figures appear at the end of this report on pp 51-106. They appear in the same order in which they are referenced.

Vertical hits were calculated as:

$$L_j = (10) \sum_{i=1}^{n} s_{ji} \qquad \text{[Eq 6]}$$

where:

L_j = hits at j^{th} level
$j = 1$ to 11
i = individual sampling points
s_{ji} = 1 for presence
= 0 for absence
n = number of sampling points (12).

Vegetation volume (VEGVOL) was calculated as:

$$VEGVOL = (10) \sum_{j=1}^{m} \sum_{i=1}^{n} s_{ji} \qquad \text{[Eq 7]}$$

where:

m = number of vertical levels (11).

Horizontal heterogeneity (HZHW) was calculated as:

$$HZHW = (10) \left[X_q(max) - X_q(min)\right] (SD_W/\bar{X}_W) \qquad \text{[Eq 8]}$$

where:

W = within quadrat

X_q = VEGVOL in transect q

SD = standard deviation.

$\bar{X}_W = (1/4) \sum_{q=1}^{4} X_q$

Vertical heterogeneity (VHW) was calculated as:

$$VHW = (10) \sum_{j=1}^{m} 1/p_j^2 \qquad \text{[Eq 9]}$$

where:

p_j = proportion of hits in j^{th} layer
$p_j = L_j/VEGVOL$

Ground cover (GC) was calculated as:

$$GC = (5) \sum_{i=1}^{20} s_{G_i} \qquad \text{[Eq 10]}$$

Canopy cover (CC) was calculated as:

$$CC = (5) \sum_{i=1}^{20} {}^{s}C_i \qquad \text{[Eq 11]}$$

Level ground (TOP) was calculated as:

$$TOP = (5) \sum_{i=1}^{20} {}^{s}T_i \qquad \text{[Eq 12]}$$

where:

${}^{s}T_i$ = 1 for hit on level ground
${}^{s}T_i$ = 0 for hit on slope.

Aquatic habitat (AQUA) was calculated as:

$$AQUA = (5) \sum_{i=1}^{20} {}^{s}A_i \qquad \text{[Eq 13]}$$

where:

${}^{s}A_i$ = 1 for hit on aquatic site
${}^{s}A_i$ = 0 for hit on terrestrial site.

Shrub density and heterogeneity (Figure 3) were estimated by walking along each of the four perpendicular transects with the arms fully extended perpendicular to the body. The total number of vertical woody stems < 3 cm diameter breast height (dbh) transected by the arms was recorded for each of the four transects. The intersections were at a height of 1.5 m. The closest shrub (woody vegetation with stem diameter < 3 cm *and* height $\geq$ 1 m) in each quarter from locus C (Figure 3) was identified; its mean crown diameter at maximum horizontal projection and its distance from C ($\leq$ 13 m) were recorded.

Shrub density (SHRD) was calculated as:

$$SHRD = (1000/92.56) \sum_{w=1}^{4} S_w \qquad \text{[Eq 14]}$$

where:

S_w = number of woody stems intersected in transect w.

Shrub heterogeneity (SHRH) was calculated as:

$$SHRH = (10) \left[S_w(\max) - S_w(\min)\right] (SD_W/\bar{S}_W) \qquad \text{[Eq 15]}$$

where:

$\bar{S}_W = (1/4) \sum_{w=1}^{4} S_w$

W = within quadrat

SD = standard deviation

Shrub cover (SHRC) was calculated as:

$$SHRC = \pi(\bar{y}_w/2)^2 \text{ (SHRD)} \qquad \text{[Eq 16]}$$

where:

$\bar{y}_w$ = mean shrub diameter

$$\bar{y}_w = (1/16) \sum_{k=1}^{16} y_k$$

Within the 26-m circular quadrat, the height of the dominant or most abundant vegetation layer (CANAV) as well as the highest vertical projection of vegetation (CANMX) were measured with the 7.5-m telescopic measuring rod or an accurately calibrated TOPCON clinometer/range finder.

Each tree whose bole center fell within the 26-m circular quadrat (dbh $\geq$ 3 cm) was identified and placed into one of 10-dbh categories (see Appendix E). The basal area categories were measured on a dbh-calibrated meter stick, but with experience, trees could be classified accurately from distances of up to 13 m.

Total basal area of trees (TREMAS) was calculated as:

$$TREMAS = (\pi/10^4) \sum_{b=1}^{10} N_b \, (\bar{d}_b/2)^2 \qquad \text{[Eq 17]}$$

where:

N_b = total number of trees of size category b

$\bar{d}_b$ = mean dbh of size category b.

Basal area of snags (DDEA) was calculated as:

$$DDEA = (\pi/10^4) \sum_{b=2}^{10} N_{b_s} \, (\bar{d}_b/2)^2 \qquad \text{[Eq 18]}$$

where:

N_{b_s} = total number of snags of size category b.

The basal area of a specific tree species or tree guild was calculated in the same manner as TREMAS, except that N_{b_x} was equal to the number of trees of species (or guild) x of size category b. Appendix E lists all 44 Illinois habitat variables. Appendix F provides a complete list of all tree species and guild memberships, and scientific names. All variables were transferred as $V_x = \log_e (V_x + 1)$.

Illinois habitat variables were grouped into four subsets for analysis.

1. The "structural" subset consists of variables defining the presence/absence or distribution of vegetation in a three-dimensional grid superimposed on each quadrat, including the aquatic and topographic variables.

2. The "typical" subset consists of variables generally seen in the literature and that researchers use consistently to quantify bird/habitat relationships: ground and canopy covers, measures of shrub abundance and

patchiness, canopy heights, species richness of shrubs and trees, and the densities of various size classes of trees.

3. The "typical + trees" subset is subset number 2 (discussed above) with the addition of variables quantifying basal areas of tree species or guilds (e.g., the addition of floristics components).

4. The subset "all" represents the use of all variables.

Data Analysis

The SAS (87) statistical package and the IBM mainframe computer were used for all multivariate analyses. Principal components solutions were obtained from Pearson product-moment correlation matrices, and were varimax rotated. Crucial to the principal components analysis was the extraction of only four principal components (PCs). Chapter 5 indicates that PC 1 is extremely important; not only is its ecological relevance critical, but its respective eigenvalue explains the largest portion of sample variance. Generally, PC 2 and/or PC 3 may also be important and interpretable, describing other ecological features unrelated (or indirectly related) to the major environmental feature defined by PC 1. Principal components higher than the first three or four are generally not interpretable, and cumulatively explain a very low portion of total sample variance. This is particularly true with environmental data sets.

The possible number of extractable principal components is, in principle, equal to the number of variables in the data set, but the higher-order components may have eigenvalues very close to zero. Chapter 5 shows that the use of varimax rotation enhances the ecological interpretation of PC 2, PC 3, and sometimes PC 4. However, the structure of the rotated factor pattern matrix depends strongly on the number of principal components transformed by the rotation. If principal components are extracted from an n variable data set, the n x n factor pattern matrix is obtained. Multiplying this matrix with an n x n transformation matrix results in the n x n orthogonally rotated factor pattern matrix. Using the above routine, the loadings on the first four rotated principal components are different from those obtained if the n x 4 factor pattern matrix is used. (The n x 4 matrix uses only the first four principal components.) The reason for the discrepancy can be attributed to the structure of the transformation matrix:

$$\mathbf{B} = \mathbf{AT} \qquad \text{[Eq 19]}$$

where:

$\mathbf{A}$ = the $n \times \ell$ factor pattern matrix
$\mathbf{T}$ = the $\ell \times \ell$ transformation matrix
$\mathbf{B}$ = the $n \times \ell$ rotated factor pattern matrix
n = number of variables
ℓ = number of principal components.

As the number of principal components that go into the rotated solution increases, the size of the $\ell \times \ell$ transformation matrix increases. This matrix is constructed from systematic, successive orthogonal rotations (transformations) such that each axis is rotated with every other axis only once. Therefore, the final complete orthogonal rotation is actually a product of

many simpler orthogonal rotations. Because of orthogonality, the elements of **T** are subjected to the following important constraint:

$$\sum_{k=1}^{\ell} \lambda_{ki}\lambda_{kj} = d_{ij} \quad (i,j=1,2,\ldots\ell;\ i \leq j) \qquad \text{[Eq 20]}$$

$$d_{ij} = 1 \quad \text{when } i=j$$
$$d_{ij} = 0 \quad \text{when } i \neq j$$

The number of such conditions is $\ell(\ell + 1)/2$. Since there are ℓ^2 possible parameters in matrix **T**, there remain $\ell(\ell - 1)/2$ degrees of freedom. The construction of **T** by successive orthogonal transformations with the above constraint inherently reduces factor loadings, obscuring the relationships between the rotated solution and the original variables. When ℓ is small, the diagonal elements of the transformation matrix have high loadings (e.g., 0.6, 0.7, 0.8, etc.); however, as ℓ increases, diagonal elements have low and inconsistent values. In other words, with increasing ℓ (number of principal components), the interpretation of the rotated solution with respect to the original variables becomes increasingly confusing.

A common error is to obtain all possible principal components from a data set, rotate their solution, and subsequently interpret the principal components analysis by retaining the first few principal components. For this study, the n x 4 factor pattern matrix has been used to obtain varimax rotated solutions; this has optimized the ecological relevance and interpretation of principal components analysis. Similarly, canonical analysis of discriminance was performed using only the first four canonical variates.

The clustering algorithm used was the hierarchical agglomerative unweighted pair-group method using arithmetic averages (UPGMA), which is also called average linkage clustering. This method was developed by Sokal and Michener (95). Squared Euclidean distance was used to construct the resemblance matrix. Habitat variables were transformed (Appendix C) and then standardized to Z scores (the means are 0 and the variances are 1) before analysis. Principal component and canonical variate scores were not transformed or standardized. This methodology was found to be robust and effective with a wide diversity of databases and consistently superior to other approaches. The dendrograms produced by unweighted clustering have higher cophenetic correlations with resemblance matrices than dendrograms produced by weighted clustering (93). Sneath and Sokal (94), Romesburg (84), and Seber (90) also strongly recommend the UPGMA. A dendrogram scale is not shown in any cluster analysis for this study, since only the relative relationships of the nodes are of primary concern.

5 DATA ANALYSIS RESULTS

Theoretical Foundation for Environmental Gradients

A series of m study sites, H_h, is identified in which the magnitude of h reflects the relative degree of environmental impact. Thus, H_1 is a control site, and H_m is the most impacted. A suite of p habitat (environmental) variables, X_j, is obtained with total sampling units $\sum_{h=1}^{m} N_h$. A sampling unit is a transect or quadrat of any size or shape within which X_j's are measured. Values of X_j are sample estimates of population parameters: means, standard deviations, coefficients of variation, etc. The N x p data matrix for site H_h is :

$$X_h = \begin{matrix} X_{11h} & X_{21h} & \cdots\cdots & X_{p1h} \\ X_{12h} & X_{22h} & \cdots\cdots & X_{p2h} \\ \vdots & \vdots & & \vdots \\ X_{1N_h h} & X_{2N_h h} & \cdots\cdots & X_{pN_h h} \end{matrix} \qquad [Eq\ 21]$$

The maximum-likelihood estimate of mean habitat vector u_h in the h^{th} site is the sample mean vector $\bar{\mathbf{x}}_h$ of the h^{th} site. Each habitat site is represented in p-dimensional habitat space by the mean habitat vector $\bar{\mathbf{x}}_h$.

$$\bar{\mathbf{x}}_h = 1/N \sum_{i=1}^{N} X_{1ih},\ 1/N \sum_{i=1}^{N} X_{2ih}, \cdots\ 1/N \sum_{i=1}^{N} X_{pih} \qquad [Eq\ 22]$$

$$\bar{\mathbf{x}}_h = [\bar{X}_{1h}, \bar{X}_{2h}, \cdots\cdots \bar{X}_{ph}]$$

By the definition used here, the *major* difference between the study sites is attributable to environmental impacts. (There are generally other differences as well.) Therefore, the loci of study sites in p-dimensional habitat space will be ellipsoidal, with the major axis corresponding to the habitat variable closely associated with characterizing the environmental impact. This would be the simple case if all habitat variables were independent. However, many environmental variables describing habitat structure are highly correlated. Although statistically a problem, this is an advantage for the ecologist, since more than one measure is available for quantifying habitat attributes such as disturbance. The realistic problem becomes to identify which habitat variables, and especially their relative contributions, express the greatest variance in the p-dimensional dispersion. Solving this problem will provide the analytical tool to quantify ecological attributes along this environmental impact gradient.

A necessary background is establishing the equivalence of coordinate system rotations and linear transformations. This is conveniently demonstrated for the two-dimensional case but, of course, has no graphical analogy when the dimensions exceed three. Examining the axis rotation in Figure 4, where the original axes are A_1 and A_2 and a rotation of θ degrees produces new axes B_1 and B_2, ($\theta_{\alpha\beta}$ defines the angle between the original axis α and a rotated axis β).

$$\theta = \theta_{11} = \theta_{22} = \theta_c = \theta_d \qquad \text{(from plane geometry)}$$

$$\cos \theta_c = b_1/(a_1 + z_1)$$

$$\tan \theta_{22} = z_1/a_2 \qquad z_1 = a_2 \sin \theta/\cos \theta$$

$$\therefore \quad b_1 = (\cos \theta)a_1 + (\sin \theta)a_2 \qquad \text{[Eq 23a]}$$

$$\cos \theta_d = b_2/z_2$$

$$\tan \theta_{11} = (a_2 - z_2)/a_1 \qquad z_2 = a_2 - a_1 \sin \theta/\cos \theta$$

$$\therefore \quad b_2 = (-\sin \theta)a_1 + (\cos \theta)a_2 \qquad \text{[Eq 23b]}$$

The original coordinates of Z were (a_1,a_2). The new coordinates of Z under rotation θ are (b_1,b_2) or, in terms of the original coordinates:

$$[(a_1 \cos \theta + a_2 \sin \theta), (-a_1 \sin \theta + a_2 \cos \theta)]$$

Two important factors should be noted:

1. Development of the new coordinates produced equations where new variables were defined that are linear combinations of the original variables; mathematically, this is a linear transformation.

2. If the locus Z represented the apex of mean vector $\bar{\mathbf{x}}$, neither the location of $\bar{\mathbf{x}}$, or more importantly, the dispersion of the raw data $[X_1, X_2, \ldots . X_p]$ (a_1, a_2 *in the two-dimensional case*) around $\bar{\mathbf{x}}$, or the loci of $\mathbf{x}_j$'s relative to one another are affected by the axis rotation (linear transformation).

$$\cos^2 \theta + \sin^2 \theta = 1 \qquad \text{(from trigonometric identities)}$$

Therefore, in any linear combination, such as Eq 23, or more generally:

$$Y = n_1 X_1 + n_2 X_2 \qquad \text{[Eq 24a]}$$

$$\text{if} \quad n_1^2 + n_2^2 = 1 \qquad \text{[Eq 24b]}$$

there exists an axis rotated by θ degrees where $n_1 = \cos \theta$ and $n_2 = \sin \theta$.

Also note from Eq 23 that:

$$(\cos \theta)(-\sin \theta) + (\sin \theta)(\cos \theta) = 0 \qquad \text{[Eq 25a]}$$

or more generally, if:

$$Y_1 = n_{11}X_1 + n_{21}X_2 \qquad Y_2 = n_{12}X_1 + n_{22}X_2 \qquad \text{[Eq 25b]}$$

$$\text{and } n_{11}n_{12} + n_{21}n_{22} = 0, \qquad \text{[Eq 25c]}$$

the axes Y_1 and Y_2 are perpendicular to each other (e.g., orthogonal).

To generalize these results to higher dimensions, it is expedient to express Eqs 23a and 23b in the form of directional cosines (Figure 4).

$$\theta = \theta_{11} = \theta_{22}$$

$$\theta_{12} = \theta + 90^0$$

$$\theta_{21} = -(90_0 - \theta)$$

$$\therefore \sin \theta = \sin (\theta_{21} + 90^0) = \cos \theta_{21}$$

$$-\sin \theta = -\sin (\theta_{12} - 90^0) = \cos \theta_{12}$$

Therefore, an equivalent form for Eqs 23a and 23b becomes:

$$b_1 = (\cos \theta_{11})a_1 + (\cos \theta_{21})a_2$$

$$b_2 = (\cos \theta_{12})a_1 + (\cos \theta_{22})a_2$$

or for p-dimensions:

$$[b_1, b_2, \ldots, b_p] = [a_1, a_2, \ldots, a_p] \begin{bmatrix} \cos_{11} & \cos_{12} & \cdots & \cos_{1p} \\ \cos_{21} & \cos_{22} & \cdots & \cos_{2p} \\ \vdots & \vdots & & \vdots \\ \cos_{p1} & \cos_{p2} & \cdots & \cos_{pp} \end{bmatrix} \qquad \text{[Eq 26]}$$

$\cos_{\alpha\beta}$ can be written as $\ell_{\alpha\beta}$ for convenience.

$\ell_{\alpha\beta}$ is equivalent to: (1) the linear transformation coefficient of element $\alpha\beta$ or (2) the cosine of the rotation angle between the original axis α ($\alpha = 1,2,\ldots,p$) and the rotated axis β ($\beta = 1,2,\ldots,p$).

The equivalence of a linear transformation and rigid axis rotation is now evident.

The condition on rigid orthogonal rotation can now be specified by expanding Eqs 24 and 25 to higher dimensions.

$$\sum_{j=1}^{p} \ell_{j\alpha}\ell_{j\beta} \qquad \begin{array}{l} \alpha = 1,2, \ldots p \\ \beta = \alpha, \alpha+1, \ldots p \end{array} \qquad \text{[Eq 27]}$$

$$= 1 \text{ if } \alpha = \beta$$

$$= 0 \text{ if } \alpha \neq \beta$$

In matrix notation, Eq 25 can be written as a row vector:

$$\mathbf{b}' = \mathbf{a}'\mathbf{L} \qquad \text{[Eq 28]}$$

with **L** being the linear transformation matrix, or equivalently as a matrix:

$$\mathbf{B} = \mathbf{AL} \qquad \text{[Eq 29]}$$

for an N x p matrix

for any $b_{\delta i} = \ell_{1\delta}a_{1i} + \ell_{2\delta}a_{2i} + \ldots. + \ell_{p\delta}a_{pi}$

$$\bar{b}_{\delta} = 1/N\sum_{i=1}^{N} b_{\delta i} = 1/N\sum_{i=1}^{N}(\ell_{1\delta}a_{1i} + \ell_{2\delta}a_{pi} + \ldots. + \ell_{p\delta}a_{pi})$$

$$= \ell_{i\delta}(1/N\sum_{i=1}^{N} a_{1i}) + \ell_{2\delta}(1/N\sum_{i=1}^{N} a_{2i}) + \ldots. + \ell_{p\delta}(1/N\sum_{i=1}^{N} a_{pi})$$

$$= \ell_{1\delta}\bar{a}_1 + \ell_{2\delta}\bar{a}_2 + \ldots. + \ell_{p\delta}\bar{a}_p$$

Therefore, in vector form:

$$\bar{\mathbf{b}}' = \bar{\mathbf{a}}'\mathbf{L} \qquad \text{[Eq 30]}$$

In matrix form:

$$\bar{\mathbf{B}} = \bar{\mathbf{A}}\mathbf{L} \qquad \text{[Eq 31]}$$

The constraint on Eq 27 in matrix form is:

$$\mathbf{L}'\mathbf{L} = \mathbf{I} \qquad \text{[Eq 32]}$$

where I is the identity matrix.

Eq 32 defines the transformation matrix **L** as an orthogonal square matrix; it is nonsingular since:

$$\mathbf{L'LL^{-1}} = \mathbf{IL^{-1}}$$
$$\mathbf{L'I} = \mathbf{IL^{-1}}$$
$$\mathbf{L'} = \mathbf{L^{-1}}$$

The problem of interest is to find the transformation matrix **L** that transforms the data matrix **X** into **Y** such that a rotated axis (Y_j) coincides with the major axis of the ellipsoid of raw data values (X_{1i}, X_{2i}, ... X_{pi}), where $[i = 1,2, \ldots \sum_{h=1}^{m} N_h]$.

The raw data matrix must be in the form of a dispersion matrix (e.g., sums of squares and cross-products), since there is interest in data variance (see Appendix G). In the well known univariate case:

$$\sum x^2 = \sum X^2 - \sum \bar{X}^2 \qquad \text{[Eq 33]}$$

where $\bar{X} = \sum X/N$ and $x = X - \bar{X}$.

$\sum \bar{X}^2$ is generally seen in the form $N(\bar{X})^2$.

The matrix form of Eq 33 is analogous, and the (p x N) data matrix can be represented over all habitat sites by:

$$\mathbf{A}(X) = \mathbf{X'X} - \mathbf{\bar{X}'\bar{X}} = \begin{matrix} \sum x_1^2 & \sum x_1x_2 & \cdots & \sum x_1x_j \\ \sum x_2x_1 & \sum x_2^2 & \cdots & \sum x_2x_j \\ \vdots & \vdots & & \vdots \\ \sum x_jx_1 & \sum x_jx_2 & \cdots & \sum x_j^2 \end{matrix} \qquad \text{[Eq 34]}$$

$$\text{Similarly, } \mathbf{A}(Y) = \mathbf{Y'Y} - \mathbf{\bar{Y}'\bar{Y}}. \qquad \text{[Eq 35]}$$

Substituting Eqs 29 and 31 into Eq 35:

$$\begin{aligned} \mathbf{A}(Y) &= \mathbf{(XL)'(XL)} - \mathbf{(\bar{X}L)'(\bar{X}L)} \\ &= \mathbf{L'X'(XL)} - \mathbf{L'\bar{X}'(\bar{X}L)} \\ &= \mathbf{L'(X'X)L} - \mathbf{L'(\bar{X}'\bar{X})L} \\ &= \mathbf{L'(X'X - \bar{X}'\bar{X})L} \end{aligned}$$

From Eq 34:

$$A(Y) = L'A(X)L \qquad \text{[Eq 36]}$$

As discussed above, we are interested in the transformation:

$$Y_j = \ell_{ij} X_1 \ell_{2j} X_2 + \ldots + \ell_{pj} x_p.$$

The dispersion of Y is given by $\sum y^2$. From Eq 36:

$$\sum y^2 = v'A(X)v \qquad \text{[Eq 37]}$$

where **v** is the vector $\begin{matrix} \ell_1 \\ \ell_2 \\ \cdot \\ \cdot \\ \cdot \\ \ell_p \end{matrix}$ or p x 1 transformation matrix

The problem is to construct **v** such that $\sum y^2$ is maximized. Maximizing $\sum y^2$ is meaningless without a stipulation on **v**. Otherwise $\sum y^2$ would be $+ \infty$. Since we only want to deal with rigid orthogonal rotations, Eq 32 identifies the constraints: $v'v = 1$ for a vector. Maximizing a function with an associated constraint is done using the Lagrange multiplier (λ).

$$G(x,y,z) = F(x,y,z) - \lambda\phi(x,y,z) \qquad \text{[Eq 38]}$$

F(x,y,z) can be maximized if:

$$\phi(x,y,z) = 0$$

and

$$\frac{\partial G}{\partial x} = 0 \quad \frac{\partial G}{\partial y} = 0 \quad \frac{\partial G}{\partial z} = 0.$$

In this case, $G(v) = F(v) - \lambda\phi(v)$ (see ref. 4).

$$\frac{\partial G}{\partial v} = 2A(X)v - 2\lambda Iv$$

when

$$\frac{\partial G}{\partial v} = 0$$

$$|A(X) - \lambda I|v = 0 \quad \lambda \text{ a scalar} \qquad \text{[Eq 39]}$$

If $|A(X) - \lambda I|$ is nonsingular, a possible solution to Eq 37 is trivial (e.g., the null vector):

$$v = |A(X) - \lambda I|^{-1}[0]$$

Therefore, $[\mathbf{A}(X) - \lambda I]$ must be singular for a nontrivial solution; mathematically, its determinant must vanish.

$$|\mathbf{A}(X) - \lambda I| = 0 \qquad \text{[Eq 40]}$$

This is the well known "characteristic equation," with p characteristic roots λ_j, or eigenvalues of the matrix $\mathbf{A}(X)$; p equals the rank of $\mathbf{A}(X)$. Since $\mathbf{A}(X)$ is a symmetric matrix, all eigenvalues are real. Furthermore, if $\mathbf{A}(X)$ is nonsingular, no eigenvalue equals zero. From Eq 37, each eigenvalue λ_j has associated with it an eigenvector $\mathbf{v}_j$. (See Appendix A.)

Solving Eqs 39 and 40 is routine; a pertinent discussion is available in any matrix algebra text, such as ref. 89.

Dividing Eq 37 by N (technically N-1 for an unbiased estimate), and specifying Y:

$$\text{var } Y_j = \mathbf{v}_j'\mathbf{S}(X)\mathbf{v}_j \qquad \mathbf{S}(X) \text{ is a covariance matrix}$$

From Eq 39:

$$\mathbf{S}(X)\mathbf{v}_j = \lambda_j\mathbf{v}_j$$

$$\text{var } Y_j = \mathbf{v}_j'\lambda_j\mathbf{v}_j \qquad \text{since } \mathbf{v}'\mathbf{v} = 1 \text{ and } \lambda_j \text{ is a scalar}$$

$$\therefore \text{ var } Y_j = \lambda_j \qquad \text{[Eq 41]}$$

Eigenvalue λ_j is identified as the *variance* of the j^{th} transformed variable, or equivalently, the j^{th} linear combination of original data variables $(X_{1j}, X_{2j}, \ldots\ldots X_{pj})$, where $Y_j = \ell_{1j}X_1 + \ell_{2j}X_2 + \ldots\ldots + \ell_{pj}X_p$.

Since we maximized variable Y in Eq 37 using Eq 38, the first solution variable $Y_1 = \lambda_1$ is a maximum. Y_1 is therefore the desired transformed variable, and equivalently, Y_1 is the major axis of the ellipsoid of $\sum_{h=1}^{m} N_h$ points in p-dimensional space of habitat variables. We can solve for any $Y_{1ih} = \mathbf{x}_{ih}\mathbf{v}_1$. For each site H_h, a mean score is calculated for Y_1 from the raw habitat variables X_{jih}.

$$\bar{Y}_{1h} = (1/N_h)\sum_{i=1}^{N_h} \mathbf{x}_{ih}\mathbf{v}_1 \qquad \text{[Eq 42]}$$

Y_1 can be identified as the environment impact gradient. Also:

$$\sum \lambda_j = \text{tr } \mathbf{A}(X)$$

(tr is the trace or sum of the diagonal elements of a matrix.) Since in a covariance matrix the diagonal elements are variances:

$$\sum \lambda_j = \text{tr } \mathbf{S}(X) = \sum_{j=1}^{p} \text{var } X_j \qquad \text{[Eq 43]}$$

From Eqs 41 and 43:

$$\sum \text{var } Y_j = \sum \lambda_j = \sum \text{var } X_j \qquad \text{[Eq 44]}$$

The sum of eigenvalues represents not only the total variance of the transformed variables, but also the total variance in the original data when the data is in the form of a covariance matrix. The proportion of variance explained by the first transformed variable is $\lambda_1 / \sum \text{var } X_j$.

A reader familiar with multivariate methods would recognize the above results as a principal components solution, with Y_1 being PC 1. Actually, p-orthogonal principal components can be extracted from an N x p data matrix, with each successive principal component explaining a decreasing proportion of total sample variance.

Any two eigenvectors, $\mathbf{v}_\alpha$ and $\mathbf{v}_\beta$ ($\alpha \neq \beta$) are independent or orthogonal (e.g., $\mathbf{v}'_\alpha \mathbf{v}_\beta = 0$).

$$\mathbf{A}(X)\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha \qquad \text{[Eq 45]}$$

$$\mathbf{A}(X)\mathbf{v}_\beta = \lambda_\beta \mathbf{v}_\beta \qquad \text{[Eq 46]}$$

$$[\mathbf{A}(X)\mathbf{v}_\alpha]' = [\lambda_\alpha \mathbf{v}_\alpha]'$$

$\mathbf{A}'(X) = \mathbf{A}(X)$ (SSCP as well as covariance or correlation matrices are symmetric.)

$$\therefore \mathbf{v}'_\alpha \mathbf{A}(X) = \lambda_\alpha \mathbf{v}_\alpha' \qquad \text{[Eq 47]}$$

From Eq 46:

$$\mathbf{v}'_\alpha \mathbf{A}(X)\mathbf{v}_\beta = \lambda_\beta \mathbf{v}'_\alpha \mathbf{v}_\beta$$

Substituting Eq 47 into Eq 46:

$$\lambda_\alpha \mathbf{v}'_\alpha \mathbf{v}_\beta = \lambda_\beta \mathbf{v}'_\alpha \mathbf{v}_\beta$$

$$\mathbf{v}'_\alpha \mathbf{v}_\beta (\lambda_\alpha - \lambda_\beta) = 0$$

Since $\lambda_\alpha \neq \lambda_\beta$:

$$\mathbf{v}'_\alpha \mathbf{v}_\beta = 0$$

Similarly, from Eq 39:

$$\mathbf{S}(X)\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha \qquad \text{[Eq 48]}$$

Note that λ_α here does not have the same values as λ_α in Eq 45.

In a complete matrix form, Eq 48 is written:

$$\mathbf{S}(X)\mathbf{V} = \mathbf{\Lambda V}$$

V is the matrix whose successive columns are $\mathbf{v}_1$, $\mathbf{v}_2$ ..... $\mathbf{v}_p$.

$$\mathbf{\Lambda} \text{ is simply } \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{bmatrix}$$

$\mathbf{\Lambda}$ is the covariance matrix of the transformed variables Y_1, Y_2, ...Y_p. This matrix shows that the variance of each Y_α is λ_α and that the covariance or correlation between Y_α and Y_β is 0, $\alpha \neq 0$ (i.e., they are independent).

By their very nature, habitat variables represent the measurement of a wide variety of environmental attributes. Obviously, variables will be measured on different scales: dimensions (linear as well as exponential functions), weights, degrees or radians, time, temperature, proportions or percentage, counts, variances or coefficients of variation, dimensionless indices, etc. However, they are all interval or ratio measures and are desirable because of their statistical properties. However, it may be more appropriate to quantify some (sometimes all) variables on an ordinal ranking. An example would be the arithmetic rating scale of 1 to 10, or even a geometric series, such as an exponential or log normal scale. Another important scale in collecting ecological data is presence/absence, where data values are either 1 or 0. Not only are all these scales vastly different, but even with variables of the same measure, the relative magnitudes and/or variances may vary considerably among environmental attributes. Generally, the relative magnitudes of means and variances are positively correlated. Following are some common examples of habitat variables of different scale: shrub height or diameter (m), shrub cover (m^2), shrub volume (m^3), shrub density (individuals/ha), substrate particle size (mm), distance from similar habitats (km), primary production (g/m^2), ground cover (%), slope aspect (degrees), shrub heterogeneity (coefficient of variation), vertical vegetation heterogeneity (a diversity index), relative soil moisture (1 to 5), and presence/absence of shrub species (1 or 0).

Another very important consideration of scaling is that it is of interest to identify high *relative* variance components that are closely associated with the environmental gradient, rather than high variance per se.

A simple way to "standardize" all environmental variables, regardless of scale or measure, is to construct standard scores for the variables. Therefore, for each habitat variable, the mean is 0 and the variance is 1. The standardized score $Z_{\alpha i}$ for the i^{th} sample of variable α is given by:

$$Z_{\alpha i} = \frac{X_{\alpha i} - \bar{X}_{\alpha}}{SD_{\alpha}} \qquad \text{[Eq 49]}$$

In other words, the new units of the habitat variables are really deviations from their respective means per unit of standard deviation (or variance). In terms of Z scores, the covariance matrix becomes the correlation matrix, since the diagonal elements:

$$(1/N) \sum_{i=1}^{N} (X_{\alpha i} - \bar{X}_{\alpha})^2 = 1 \quad \text{for all } \alpha,$$

and the rest of the elements:

$$(1/N) \sum_{i=1}^{N} (X_{\alpha i} - \bar{X}_{\alpha})(X_{\beta i} - \bar{X}_{\beta}), \ \alpha \neq \beta,$$

are equivalent to $r_{\alpha\beta}$, the correlation coefficient between X_{α} and X_{β}.

The author feels that variable standardization justifies use of the correlation matrix instead of the covariance matrix when dealing with habitat variables, even though the covariance matrix has better statistical properties (3, 44, 71, 39, 17). The two main problems are: (1) distributional properties and sampling theory (including hypotheses testing and calculation of confidence intervals) of transformed variables extracted from correlation matrices are much more complex than those derived from covariance matrices; and (2) maximizing the variance of standardized scores intuitively appears ambiguous. A careful assessment of the arguments does not preclude the necessity for variable standardization, especially when primary interest is in the first transformed variable (e.g., PC 1). The author's experience has been that when variables scales are heterogeneous, the correlation matrix gives results that are ecologically more interpretable than an eigenanalysis using a covariance matrix.

PCA can also be performed on the dispersion matrix (Appendix G). This is known as noncentered PCA (74, 113, 11). Noncentered PCA has advantages in classification or describing compositional variation that include discontinuities, but is subject to strong distortion (113).

This discussion has centered on finding Y_1, which represents the environmental impact gradient. Actually, an eigenanalysis (or principal components analysis) extracts p-transformed variables (principal components): Y_1, Y_2,.....Y_p, representing linear combinations of the original p-habitat variables. Some of these transformed variables may be ecologically interpretable, and since they are orthogonal, they would represent ecological gradients that are uncorrelated with habitat disturbance. The problem is that the p extracted transformed variables, each containing unique linear combinations of correlated habitat variables, are difficult to interpret realistically. A

similar analogy was seen with the original scatter of $\sum_{h=1}^{m} N_h$ points in p-dimensional space. In a similar fashion, an orthogonal rotation of axes can be envisioned that produces closer agreement between the transformed variables (also called factors) and the original variables. This is referred to as "simplifying factor structure." Thurston (103, 104, 105) first developed qualitative guidelines or principles for parsimony, or simple structure; a lucid discussion can be found on pages 94 through 99 of ref. 43.

The first criteria for analytical rotations were developed by Carroll (12); an extensive review is provided by Harman (43). Ferguson's (30) approach to mathematically defining parsimony for the rotational solution can serve as a simple explanation for the numerous analytical procedures that have been developed, since most of the solutions are closely related. Ferguson reasoned that the simplest description of a point in space upon rotation of orthogonal axes is when one of the axes passes through the point. As one of the axes is rotated toward the point, the product of the two coordinates becomes smaller. He therefore suggested that some function of the sum of products of coordinates be used as a parsimonious description of these points. Since coordinates take on negative as well as positive values, he proposed the sum of squares of coordinate products instead of the simple sum. This is exactly analogous to the familiar calculation of sum of squares as a measure of dispersion around the mean. Analytically, this is:

$$\sum_{j=1}^{P} \sum_{q<r=1}^{R} (a_{jq} a_{jr})^2$$

with R(R-1)/2 sums of p pairs of coordinates. Parsimonious solutions to functions of this general structure reduce to simplifying the rows *or* columns of the factor matrix derived from the eigenanalysis or principal components solution. The factor matrix contains the "loadings" or the Pearson product-moment correlation coefficients of the principal components (transformed variables) with the original variables. PC 1 is in the first column, PC 2 is in the second, etc.

The quartimax criterion (72) simplifies the *rows* by finding orthogonal transformations that maximize the variance in the distribution of squared factor loadings. Carroll (12) and Saunders (88) proposed similar solutions for row simplification. Simplifying the rows of a factor matrix has the effect of associating the variables with fewer factors. Thus, parsimony is achieved by producing a few generalized factors. In practice, the first principal component (often PC 2 as well) has many variables associated with it and is therefore difficult to interpret ecologically.

The varimax criterion (59) is an effective orthogonal rotation that enhances the ecological interpretation of principal components $\geq$ 2. Essentially, the varimax method simplifies the *columns* of the factor matrix. This has the effect of making the correlations in each column approach 1 or 0, and therefore minimizes the number of moderate loadings on each principal component. The tendency is for a given principal component to be strongly associated with a smaller set of original variables and therefore easier to interpret. The varimax method is particularly effective for enhancing the interpretation of PC 2 and PC 3, and sometimes PC 4. Principal components

greater than four are not usually interpretable because they explain such an insignificant portion of sample variance.

In summary, the first orthogonal rotation identified as Y_1 represents the linear combination of original habitat variables that explains the greatest component of sample variance. Subsequent orthogonal rotations (under the varimax criteria) simplify the columns of the factor matrix, and therefore usually enhance the ecological interpretation of Y_2, Y_3, and sometimes Y_4.

Habitat Variables

Mojave Desert Sites (Low Ratio of Number of Sampling Units to Number of Variables)

Canonical Analysis of Discriminance. Figures 5 through 10 show the relative discriminating ability of habitat-variable subsets to quantitatively classify the five study sites of the central Mojave Desert. Tables 1 through 7 give the Pearson correlation coefficients and their significance for habitat variables with the first four canonical variates. These data represent seven logical assortments of habitat-variable subsets. In reality, more than 50 variable subsets were analyzed. Ordination of study sites in the canonical space of habitat variables depended strongly on the specific choice of variables. The complete variable set could not be analyzed, since the number of variables for canonical analysis is limited to the number of sampling units. Figure 5 and Table 1 show the analysis using the 18 variables from the original data set that passed the tolerance test. (This ensures that the rows and columns of the data matrix are linearly independent; therefore, the data matrix is nonsingular and has an inverse--a prerequisite for canonical analysis.) CV 1 separates study sites on the basis of environmental characteristics of valleys contrasted to alluvials/high desert, and CV 2 separates these sites on the basis of habitat disturbance. The gradient along CV 2 is not very clear (Table 1). Note the location of site M in Figure 5.

The analysis using only 10 "basic" habitat variables permits a clearer ecological interpretation (see Figure 6 and Table 2). CV 1 separates study sites by contrasting dense shrub and/or forb cover with sandy substrates--an environmental impact gradient. CV 2 separates the valley control from the other sites on the basis of the valley control's high burroweed and grass cover and its coarse sand substrates.

The use of shrub height classes along with the "basic" variables (Figure 7 and Table 3) clearly separate the impacted sites from the controls on CV 2. CV 1 effectively separates alluvials and valleys, contrasting gravel substrates and dense low shrub cover with grass and coarse sand substrates, respectively. However, the desirable feature of Figure 6 is lost, CV 1 representing the environmental impact gradient.

Figure 8 and Table 4 represent the "basic" variables, with two variables added to quantify shrub size heterogeneity (variability in the diameter of creosote bush and burroweed). These two species usually account for more than 90 percent of the shrub cover in the central Mojave (64). Variability in the heights of these species correlates strongly with variability in diameter (Pearson, $r = 0.86$ for creosote bush, and 0.80 for burroweed, $P < 0.001$) so

only one of these measures is sufficient. Figure 8 amplifies the interpretation of Figure 6 by adding creosote bush heterogeneity (crushed shrubs) to sandy substrates as characteristics of disturbed Mojave Desert habitats. CV 2 contrasts grass and coarse sand substrates with gravel substrates and burroweed size heterogeneity. Therefore, CV 2 contrasts environmental characteristics of valleys with those of alluvials/high desert, respectively. Note the locations of S and M on Figure 8, and note that CV 1 portrays an environmental impact gradient. The addition of variable CL4 (cover of shrubs $\geq$ 1.5 m in height) had no effect on the canonical structure of Figure 8 (see Table 5). Shrubs of this height are important characteristics of unimpacted desert valleys (64).

The addition of variables for the mean heights of creosote bush, burroweed, and subdominant shrub species (even though these are highly correlated with their respective cover variables) helped separate the valley control from the other sites (See Figure 9 and Table 6). Note that forb and subdominant shrub cover and gravel substrate have been removed from CV 1. Since these variables characterize alluvials, they not only enhanced the separation of alluvial and valley controls on CV 1, but sharpened the alluvial-valley gradient on CV 2. Since shrubs more than 1 m tall characterized unimpacted habitats, while shrubs less than 1 m tall were typical of disturbed Mojave habitats (64), the variable L12J (density of shrubs 1.1 to < 1.5 m tall) was selected as another potential discriminating variable. The resulting variable subset produced the best discriminating subset of variables to effectively classify study sites; CV 1 clearly defines the environmental impact gradient (See Figure 10 and Table 7). Note that the alluvial control and the high desert have similar habitat structure.

Eigenanalysis. Figure 11 shows the varimax rotated principal components solution that used the habitat variable subset (Table 7) which provided the best ecological interpretation in canonical space. Table 8 gives the Pearson correlation coefficients of each habitat variable with the first four principal components (eigenvectors). Clearly, this ordination is inadequate.

Figure 12 and Table 9 show the ordination using the complete set of habitat variables. Although the ordination presents a better classification than the one shown in Figure 11, the ecological significance of some of the variables is obscure (e.g., the presence of many of the shrub species in only some vegetation transects).

By screening more than 30 variables subsets and sequentially adding variables of known importance (64), an optimal set of habitat variables was delimited. Principal components analysis was not as sensitive as CAD to variable selection. Also, a larger number of habitat variables could be included in variable subsets, since eigenanalysis solutions are not as critical to matrix row/column linear independence as is canonical analysis.

Figure 13 and Table 10 show the optimal ordination. PC 1 represents the environmental impact gradient--dense cover of creosote bush and burroweed, tall shrubs, litter accumulation, and coarse sand substrates--contrasted to high variability of creosote bush size (crushed shrubs) and sandy substrates. PC 2 is positively associated with environmental characteristics of alluvials and the high desert--cover of subdominant shrub species, high variability of burroweed size (large burroweed along washes), short creosote

bush, and gravelly and rocky substrates. PC 3 contrasts grass and density heterogeneity of creosote bush with forbs, shorter creosote bush, and density heterogeneity of all shrubs other than creosote bush. These environmental characteristics helped separate the impacted alluvial from its control.

Illinois Successional Sites (High Ratio of Number of Sampling Units to Number of Variables)

Canonical Analysis of Discriminance. The Illinois data set was organized into four subsets of habitat variables: structural, typical, typical + tree guilds, and all (see Chapter 4 and Appendix E for more details). Figures 14 through 17 show the mean canonical variate scores (centroids) of the 10 study sites on the first four canonical variates. Tables 11 through 14 give the Pearson correlation coefficients and their significance for each habitat variable with the first four canonical variates. Although there are differences among these four analyses, unlike CAD with the Mojave Desert variables, the general pattern is similar, regardless of which variable subset is used; CV 1 consistently defines the major environmental gradient (a succession gradient in this case). This "stability" is directly attributable to the high ratio of sample sizes to the number of variables, and was maintained even though the data "quality" was lower than that of the Mojave Desert example. (For example, in the Mojave Desert, each habitat parameter was estimated from larger sample sizes, and the samples were collected over a much larger area; see Chapters 1 and 4 for more details.)

The variables describing the habitat as the presence/absence and distribution of vegetation in a three-dimensional grid (Figure 14) gave very similar results as the analysis defining the habitat in terms of ground cover, shrub characteristics, and the density/distribution of tree sizes (Figure 15). Note that the ecological interpretation of CV 2 and CV 3 are reversed between these two analyses (Tables 11 and 12). This is very common in discriminant analyses. In both analyses, CV 4 does not have much discriminating power.

The addition of habitat variables identifying tree species in terms of their basal areas improves discriminating ability, effectively separating bottomland (G,H) from upland (I,J) forests, not only along CV 1 but also along CV 2 (see Figure 16 and Table 13). CV 2 also separates the older shrub habitats (D,E,F) from early succession stands (A,B,C) on the basis of the densities of medium-sized trees, including cottonwood and snags. CV 3 separates sites on the basis of small tree (dbh < 15 cm) density, particularly black locust.

The analysis employing all variables (see Figure 17 and Table 14) gives results comparable to those of the preceding analysis. CV 2 separates upland and bottomland forests, and CV 3 rather accurately ranks sites on a gradient from open habitats to increasing densities of shrubs and small trees (similar to the preceding analysis). CV 4 portrays complex relationships in subcanopy vegetation structure.

Eigenanalysis. Figures 18 through 21 show the mean principal component scores of the 10 study sites on the first four principal components (eigenvectors). Tables 15 through 18 give the Pearson correlation coefficients and their significance for each habitat variable with the first four principal components, and summarize the association of important habitat variables with each of the principal components. The general pattern of the four analyses is

similar. Groups of variables representing similar habitat components were associated consistently. PC 1 represents the main environmental gradient--a successional gradient in which early seres (A,B,C), shrublands (D,E,F) and forest (G,H,I,J) are separated.

Two study sites--C and F--merit additional discussion. Site F is a 56-year-old abandoned strip mine. Unreclaimed strip mines are characterized by an extensive ridge and valley topography. The ridge-and-valley relief of unreclaimed mines of this age is very low, since the ridges have eroded, and the valleys and depressions have accumulated silt. The troughs contain permanent water with extremely deep silt deposits (> 15 m). Local hydrology has helped maintain a high water table. These environmental conditions have not only limited tree density, but dramatically decreased the growth rates of volunteer trees on the ridges. A study of growth rings from tree borings verified extremely low growth rates over the last 20 years. Site F consisted of a mosaic of open areas and very dense shrub cover interspersed with medium to large trees and an unusally high representation of woody vines. Therefore, because of its low tree density and the presence of open habitats, this site ranked below sites D and E on the succession axis in three of the four analyses. The analysis that uses all habitat variables ranked it correctly in an age sequence. Note that Site F is located near the origin (0,0) in PC 1/PC 2 space that incorporates all variables (Figure 21).

Site C has very high densities of saplings that are < 15 cm dbh (mostly < 8 cm dbh), and therefore, has higher scores than sites A and B when these variables are emphasized (Figures 19 and 20). However, when the structural variables are considered (Figures 18 and 21), site C has lower scores than sites A and B. This has been interpreted as strong evidence that the environmental gradient PC 1 is, in reality, not directly representing habitat characteristics of age per se, but rather is tracking the complexity of vegetation structure, which includes vegetation volume as an important component. (Of course, vegetation complexity and volume are usually associated with successional age.) Except for a few ponds, site C consists almost entirely of two layers of vegetation: a ground cover of smooth brome grass (<u>Bromus inermis</u>) and a uniform-aged stand of young black locust (<u>Robinea pseudoacacia</u>). The use of all variables for the PCA clearly expresses this vegetation complexity/volume gradient (Figure 21).

Structural Variables (See Figure 18 and Table 15)

The relative relationships of the four forest sites are not easily interpretable with this subset of variables, except in the case of bottomland forests G and H along PC 4. The younger bottomland forest (G) has more openings, and therefore has both a higher patchiness and a higher density of low vegetation (1 to 3 m). Note also the location of the "shrubby sites" on PC 4.

PC 2 represents the density of "midstory" vegetation (2 to 7.5 m). Site C--the young black locust grove--ordinates strongly on PC 2. Site D--an older and more complex black locust grove and upland forests (I,J) also shows good representation of this vegetation layer. Early successional stages (A,B) occur at the other end of this gradient.

PC 3 contrasts high ground cover of vegetation on the positive axis with the presence of aquatic habitats on the negative axis.

Typical Variables (See Figure 19 and Table 16)

The first two principal components clearly ordinate the study sites into four categories: open habitats, shrub/saplings, bottomland forests, and upland forests. PC 2 represents a strong gradient of shrub/sapling density. In the case of the forest sites, "shrubs" represent the seedlings of canopy species. Note that bottomland forests and early successional seres have low scores on PC 2.

PC 3 represents the density of large trees (> 53 cm dbh). The older bottomland forest with fast-growing species, such as cottonwoods (Populus deltoides) and sycamores (Platanus occidentalis), have the highest scores on this axis. At site I, there are some large representatives of these species in the riparian zone.

PC 4 represents a gradient of high ground cover combined with a high species richness of shrubs. Sites B and C score low on this gradient because of the presence of aquatic habitats as well as a poor representation of shrub species. More than 97 percent of the shrubs at site C were black locust sprouts or seedlings, while most of the shrubs at site B were willow clumps (Salix interior) around ponds, with scattered rose (mainly Rosa multiflora, some R. rugosa), and a few seedlings of black cherry (Prunus serotina), cottonwood, sycamore, and box elder (Acer negundo).

Typical + Tree Guild Variables (See Figure 20 and Table 17)

The addition of tree guilds or species (see Appendices E and F) strongly separated bottomland and upland forests on PC 2. In comparison to the previous analysis, these forests are also more clearly separated on PC 1.

PC 3 represents high shrub cover and small tree (< 15 cm dbh) density along with high shrub heterogeneity (shrub patchiness).

PC 4 represents the basal area of black locust. This species was planted on sites C, D, and E. (Appendix D provides the ages of black locusts, taken from tree-boring data, for these sites). Overall, this analysis probably overestimates the importance of tree species compositions in ordinating study sites.

All Variables (See Figure 21 and Table 18)

Use of all the variables provides the most informative ordination, as well as the most realistic ecological interpretation. PC 1 accurately ordinates the vegetation's relative complexity/volume. PC 2, the second most important environmental gradient, represents the gradient of open habitats (negative scores) to that of dense shrub cover and small trees (positive scores).

PC 3 contrasts environmental characteristics of bottomland forests (very large trees, open subcanopy, and bottomland tree species) with characteristics of upland forests (dense subcanopy and seedlings, and upland tree species). Note the gradient of decreasing subcanopy from J-I-G-H on PC 3. A parallel decrease in vegetation complexity/density occurs along PC 1, but is not nearly as striking.

PC 4 represents a gradient of heavy ground cover (positive) versus the presence of aquatic habitats.

Classification Using Cluster Analysis

The classification results derived from cluster analysis depended on the specific combinations of habitat variables used. Figures 22 through 24 show, respectively, the results obtained by using the best subset of variables for CAD (Table 7), the best subset for PCA (Table 10), and the complete set of habitat variables (Appendix B). The use of all 58 habitat variables produced the best classification, and correctly placed most of the individual transects into their respective study sites. Also, this analysis immediately separated the impacted sites from the controls. Since the clustering algorithm spatially relates observations (transects) in terms of arithmetic distances in the n-dimensional space (axes) of n predictor (independent) variables, cluster linkage occurs relatively late when many environmental variables are used; therefore, transects are grouped only when they have many different environmental attributes in common. Note that the relative length of the linkage modes increases when more variables are used (Figures 22 through 24).

The cluster analyses (Figures 22 through 24) summarize or reveal several interesting features. The alluvial control and the high desert are environmentally very similar. Transects VC1 and VC2 are very different from transects VC3 and VC4, unless shrub species compositions are taken into account (Figure 24). VC3 and VC4 are almost parallel and located on a slight slope, whereas VC1 and VC2 parallel each other on the valley bottom. VC1 paralleled a rocky ridge at 125 to 200 m. Interestingly, VC1 resembles M2 (Figure 23). M2 is low on an alluvial, and Army training activities have made the substrate sandy. Transects S1 and S2 are very similar. These two transects are located in the most severely disturbed portion of study sites. Although not located close to one another, S3 and S6 share common environmental features. In the S study area, shallow, irregular washes were present only in these two transects. These transients not only had patches of larger shrubs, but increased shrub species richness, since washes in desert ecosystems are concentrated sources of water during showers or the rainy seasons.

Figures 25 through 28 show the cluster analysis results, using principal component scores as predictor variables. Only four eigenvectors were rotated in the varimax solution of PCA. The use of all four principal components (Figure 25) separated the high desert and the alluvial control, emphasized the uniqueness of VC1, recognized the similarity of S1 and S2, and classified S3 and S6 (the least impacted transects of the S study site) with the moderately impacted transects.

The use of the first three principal component scores (Figure 26) classified sites AC and L together, but in minor details (mainly that

relationships of VC3 - VC4 and S1 - S2) was similar to the classification obtained with four principal component scores.

Use of the first two principal component scores (Figure 27), which explain 70.1 percent of the original data variance, classified the high desert with the moderately impacted alluvial. Two transects from the moderately impacted alluvial (M2 and M3) were grouped with transects S3 and S6.

Figure 28 shows the classification using only one variable--PC 1 scores (PC 1 explained 48.2 percent of the data variance). This variable represents a linear combination of original habitat variables that define the environmental impact gradient. Initially, three groups of transects are clustered: (1) severely impacted, (2) controls, and (3) lightly impacted high desert/moderately impacted alluvial (including one AC transect). A more detailed clustering represents an excellent classification, particularly in terms of environmental disturbance. Figure 28, employing only a single variable for clustering, provides information similar to that given in Figure 24 (58 variables used). Also, the use of only PC 1 scores separates AC and L more effectively, and recognizes the close environmental similarity of S3 to S6, VC3 to VC4, and AC3 to L transects. This analysis also simplifies the interpretation of VC1 by classifying it with three other control transects. Figure 28 appears to represent the most realistic classification, despite the fact that only a single predictor variable was used.

A series of cluster analyses was also performed on varimax rotated principal components solutions using the maximum possible number of eigenvectors. Most researchers do this routinely. For these analyses, the best subset of habitat variables for PCA was used (the same variables were used for Figures 25 through 28), and the number of extracted eigenvectors equaled the number of variables (a total of 25). Figures 29 through 32 show the results of these analyses using the first principal component, then the first four, the first 10, and the first 20, respectively. Note that as the number of principal components used in the cluster analysis increases, it becomes more difficult to form clusters. In other words, each transect becomes more unique. Also, not only do relationships among transects become more obscure, but the realism and ecological interpretation of the classification is lost. Even the use of only one, or several of the major principal components does not produce a good classification. In Figure 29, the least and most impacted transects at S (S1 and S6) are grouped, transects from AC, M, and L are grouped in two different cluster sets, and transect S4 is identified as unique. In Figure 30, transects from four of the study sites (M, L, VC, AC) show ambiguous relationships.

Figures 33 through 36 show the cluster analyses obtained by using canonical variate scores. Note how rapidly all the experimental transects are correctly classified into their respective study sites. Use of either the first four, the first three, or the first two canonical variate scores gave essentially identical results. Like PC 1, CV 1 represents the environmental impact gradient, but cannot distinguish among control transects, classifying alluvial and valley controls into a single group.

6 DISCUSSION AND DATA SUMMARY

The analytical description of environmental gradients has been shown to be an eigenanlysis problem, and is a general solution of the well known principal components solution. In environmental impact assessment, this gradient defines habitat disturbance and corresponds to principal component 1. Higher-order environmental gradients *may* be extracted that represent other, possibly important, ecological attributes that are independent of environmental impacts. If they are present, PC 2 and PC 3 and sometimes PC 4 represent, respectively, in decreasing importance, these other gradients.

Principal components analysis was robust with respect to variable combinations selected, number of variables, sample sizes, and data matrix singularity (row/column linear dependencies). The first few eigenvectors were affected very little by whether the input matrix was of full rank--an important consideration when sample sizes are small. However, PCA analysis and interpretation benefit by having large ratios of sample sizes to number of variables.

PC 1 consistently represented the environmental impact gradient (Figures 11 through 13 and 18 through 21). Interestingly, the use of only one variable--PC 1 (derived from 25 original variables)--classified 23 study transects in the Mojave Desert as well as all 58 of the original habitat variables (compare Figures 28 and 24). Therefore, it is concluded that the statistical/ordination properties of PCA, along with its accepted popularity in a wide variety of disciplines, and readily available software for mainframe, mini- and microcomputers make PCA an excellent choice for environmental impact assessments and parsimonious description of habitat structure. Furthermore, it represents an analytical framework for assessing species/habitat associations along environmental gradients (63, 64).

PCA limitations have been reviewed, with strong emphasis on nonlinear relationships among environmental variables (or their subsets) and the position of study sites or sampling points along environmental gradients (100, 5, 34, 35, 48, 37, 75, 113, 49, 36, 33). Ordination techniques emphasizing nonlinear responses are being explored, and comparative evaluations will be discussed in a separate USA-CERL Technical Report.

PCA is bound by a number of statistical assumptions that are similar to univariate parametric statistical inferences (71, 4, 90). However, in most applications, PCA has been found to be robust to heteroscedasticity and nonnormality (44, 40, 41, 67, personal observation). This is an important consideration, since ecological data rarely conform to parametric assumptions. Good evidence of robustness was that PCA reproduced relatively similar ordinations, both with raw data and transformed data. Variable transformations vastly improve the fit of raw data to parametric assumptions, greatly reducing skewness, variance heterogeneity, and nonlinearity (96, 20a, 20, personal observation). Nevertheless, all data were transformed before analysis (See Appendix C and Chapter 4). Cassie and Michael (13), Whittaker (113), and Pimentel (78) suggest data transformations in PCA.

Whenever variables of different scales and measures (such as environmental or habitat variables) are incorporated into PCA, variable standardization

(e.g., Z scores) was found to produce ecologically more relevant ordinations. Therefore, correlation matrices were used for all PCA input. Since principal components are not invariant under linear transformations, the extracted eigenvectors of a correlation matrix are not equivalent to, nor can they be derived from, that of a covariance matrix. Although researchers use correlation matrices more frequently for PCA (17), covariance matrices are theoretically more desirable because they have better statistical/mathematical properties related to statistical inference, distribution theory, and interpretations (3, 44, 71, 39, 17). Also, it may be more difficult to interpret the coefficients given by the eigenvectors (90). These arguments have been assessed carefully, and the practicality and advantages of variable standardization, especially since primary interest is in the first principal component, far outweigh theoretical arguments and mathematical vigor.

The varimax criteria (59) for orthogonal rotation of principal component solutions was used in all PCA to simplify factor structure (43). This technique enhances the ecological interpretation of PC 2, PC 3, and PC 4. Only the first four principal components were used for varimax rotation, since the use of high-order transformation matrices resulted in rotated solutions for which factor pattern matrices (principal component loadings) were difficult to interpret in terms of the original predictor variable set. This was the direct consequence of the very small eigenvalues associated with higher-order principal components. (See Chapter 5 for more explicit details.)

Canonical analysis of discriminance proved very useful for exploratory data analysis and for selecting habitat variables that were useful for distinguishing study sites (Figures 10 and 14 through 17). Use of at least the first two canonical variates accurately classified all transects into their respective study sites in a desert ecosystem (Figure 35). Use of only the first canonical variate, which represented the environmental impact gradient (optimal CAD variable subset) could not distinguish between the two control study sites (Figures 36 and 10). The addition of canonical variate 2, representing a gradient of habitat characteristics that contrasted alluvials to valleys, was necessary for the complete discrimination of study sites (Figure 10). CAD was also used successfully to classify the Mojave Desert study sites on the basis of bird or small mammal community structure (63, 64).

This report has shown that the results of canonical analyses were strongly dependent on which combination of variables was selected for analysis (Figures 5 through 10). Canonical analysis with various combinations or subsets of variables was very sensitive to small sample sizes. Furthermore, discriminant analyses are very sensitive to linear dependency among variables (e.g., values of a given variable cannot be similar to linear combinations of other variables--a feature difficult to avoid with highly correlated environmental variables and small sample sizes). These difficulties of CAD are independent of the scaling, accuracy, precision, or dispersion of collected data values. However, appropriate variable transformations are beneficial, because they reduce heteroscedasticity of covariance matrices. The use of large sample sizes with respect to the number of environmental variables measured is highly recommended, since it minimizes the usual problems encountered with CAD.

Cluster analysis was not as effective as PCA and CAD in consistently and accurately classifying the Mojave Desert study transects. PCA and CAD

ordinations persistently associated individual transects with their respective study sites. This was not surprising, since these ordinations were derived by weighing the original variables according to a "recipe" that can uncover underlying patterns (if they exist). As long as relevant ecological patterns are present (e.g., impact or succession gradients), techniques such as PCA or CAD will be superior to clustering analysis. Recall that cluster analysis was most effective when using variables derived from other multivariate techniques, especially CAD. R-analysis--the clustering of variables on the basis of samples--is recommended as an excellent technique for initial exploratory data analysis, reducing data dimensionality, identifying independent variable subsets, and eliminating redundant variables.

7 CONCLUSIONS

The concept of environmental gradient analysis has been demonstrated to be a parsimonious and rigorous methodology for quantifying environmental impact assessments or successional seres. The analysis has potential for analytically describing a broad range of environmental or ecological phenomena.

The analytical description of environmental gradients has been shown to be an eigenanalysis problem, mathematically equivalent to the largest eigenvector (or first principal component) of a principal components analysis. The second or third, and sometimes the fourth, principal components may also have relevant environmental/ecological information. Specific suggestions have been made to optimize the ecological relevance and interpretation of principal components ordinations in terms of environmental gradients.

The analytical representation of an environmental gradient, itself a *single* independent (predictor) variable, had environmental information similar to that in the combination of all original 58 habitat variables that described five different study sites at Fort Irwin. The basis of this finding is that each produced a similar and accurate classification of the 23 Mojave Desert study transects.

Principal components ordinations were robust with respect to sample sizes, variable choice, number of variables, data matrix singularity, heteroscedasticity, and nonnormality. Therefore, principal components analyses represent an excellent analytical basis for environmental impact assessments and parsimonious description of habitat structure, as well as assessing species-habitat associations along environmental gradients.

Ecological ordinations with discriminant analyses have been shown to be sensitive to sample sizes, the choice of descriptive variables, and linear dependency among variables. Although preliminary variable screening using bivariate correlations and cluster analysis (R-mode) to eliminate variable redundancy is highly recommended, ecological realism is often enhanced by incorporating several highly intracorrelated variables to quantify an environmental attribute. Canonical analysis of discriminance may be a useful mechanism for choosing among variable subsets. Despite these analytical approaches for variable selection or rejection, it is ultimately the experience of the researcher and his/her familiarity with the environmental/ecological process being modeled that plays the major role is selecting the environmental variables for data analysis. Variable choice, as well as ordination sensitivity, become less significant when the ratio of sample sizes to number of variables increases. Interestingly, this property is independent of the variance, precision, or accuracy of the collected data, but is a fundamental property of the matrix mechanics inherent in multivariate analyses. Therefore, variables must be chosen with care and ecological insight, and adequate sample sizes are mandatory.

Principal components analysis, canonical analysis of discriminance and cluster analysis are all important methodologies for classifying environmental/ecological information. However, each has its limitations and shortcomings. There is no optimal recipe for all or even most environmental

situations. The specific nature of the environmental/ecological investigation, the desired research or interpretive goals, the resources available or committed to data collection or analyses, and the experience of the researcher all play major roles in ordination and classification technologies.

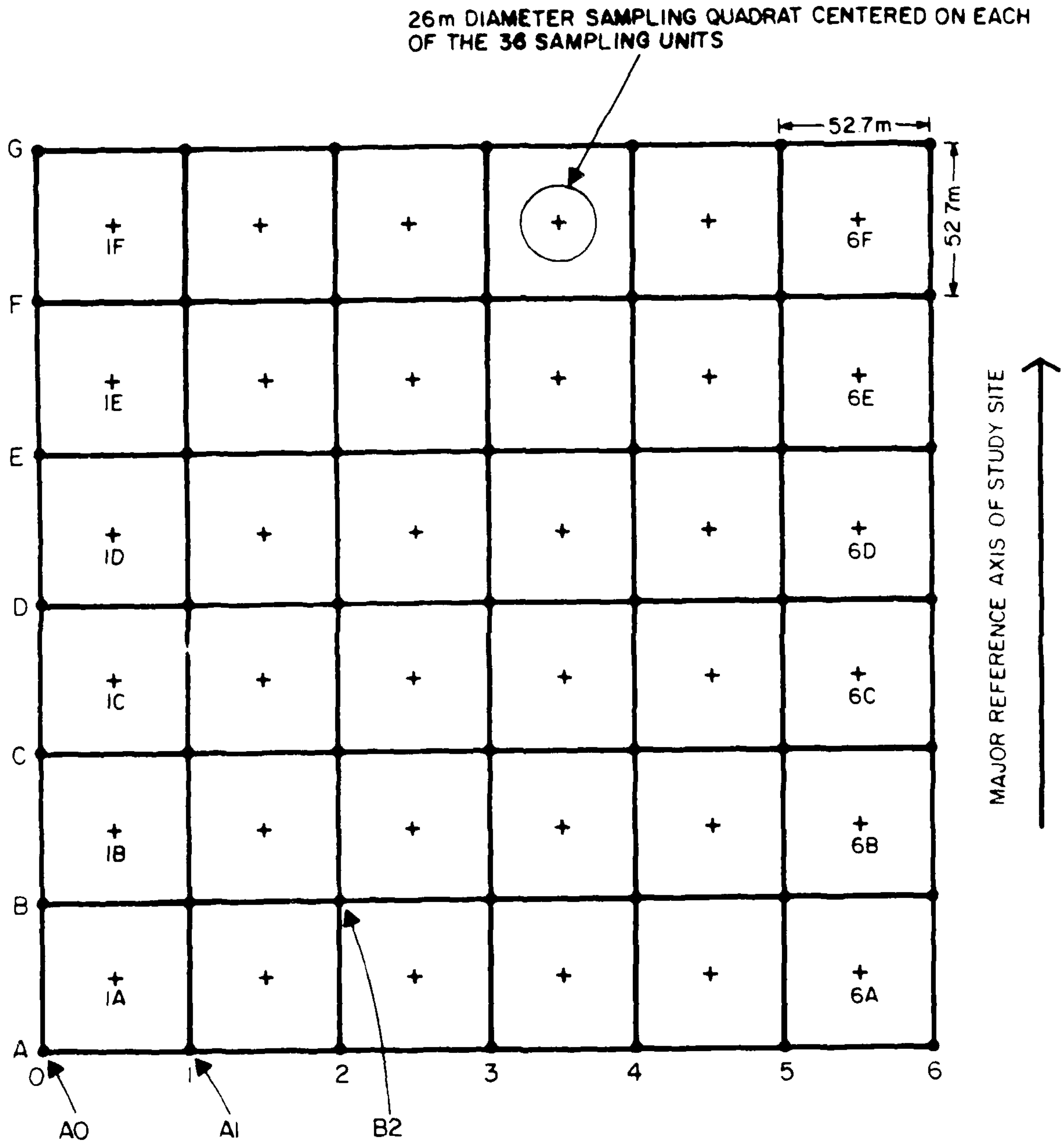


Figure 1. The Illinois 10-ha study sites gridded into thirty-six 52.7-m-square sampling units. (The corners as well as the center of each sampling unit were identified with a number/alphabet code for uniqueness. Habitat parameter estimates were obtained in the 26-m-diameter sampling quadrats.)

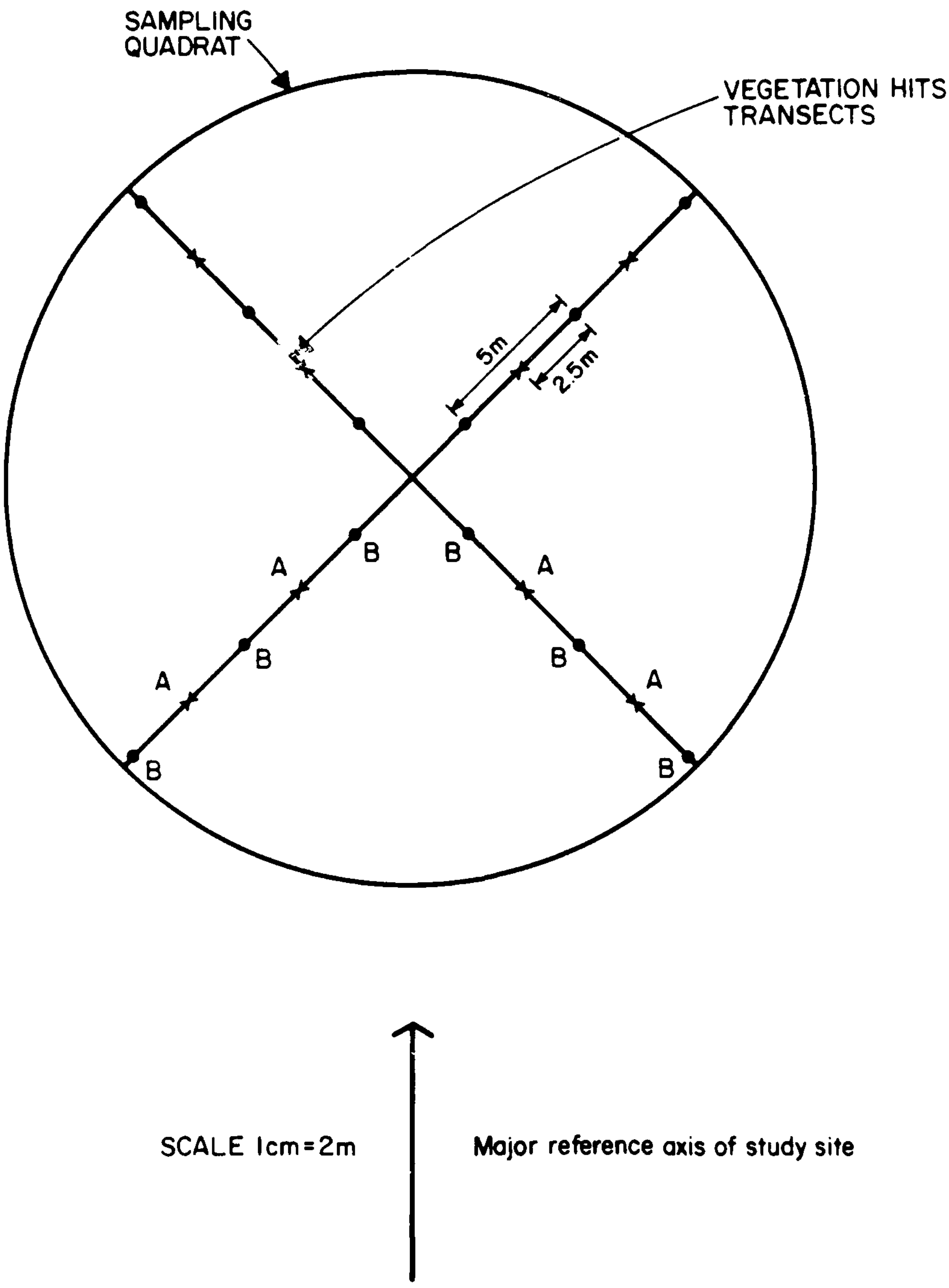


Figure 2. Schematic representation of the 26-m-diameter sampling quadrat located within each Illinois sampling unit showing the relative orientation of the vegetation "hits" transects. (At point "A," the presence/absence of ground cover, canopy cover, level topography, and aquatic habitat was recorded. At point "B," the same data were collected, with the addition of the presence/absence of vegetation in 11 vertical height categories.)

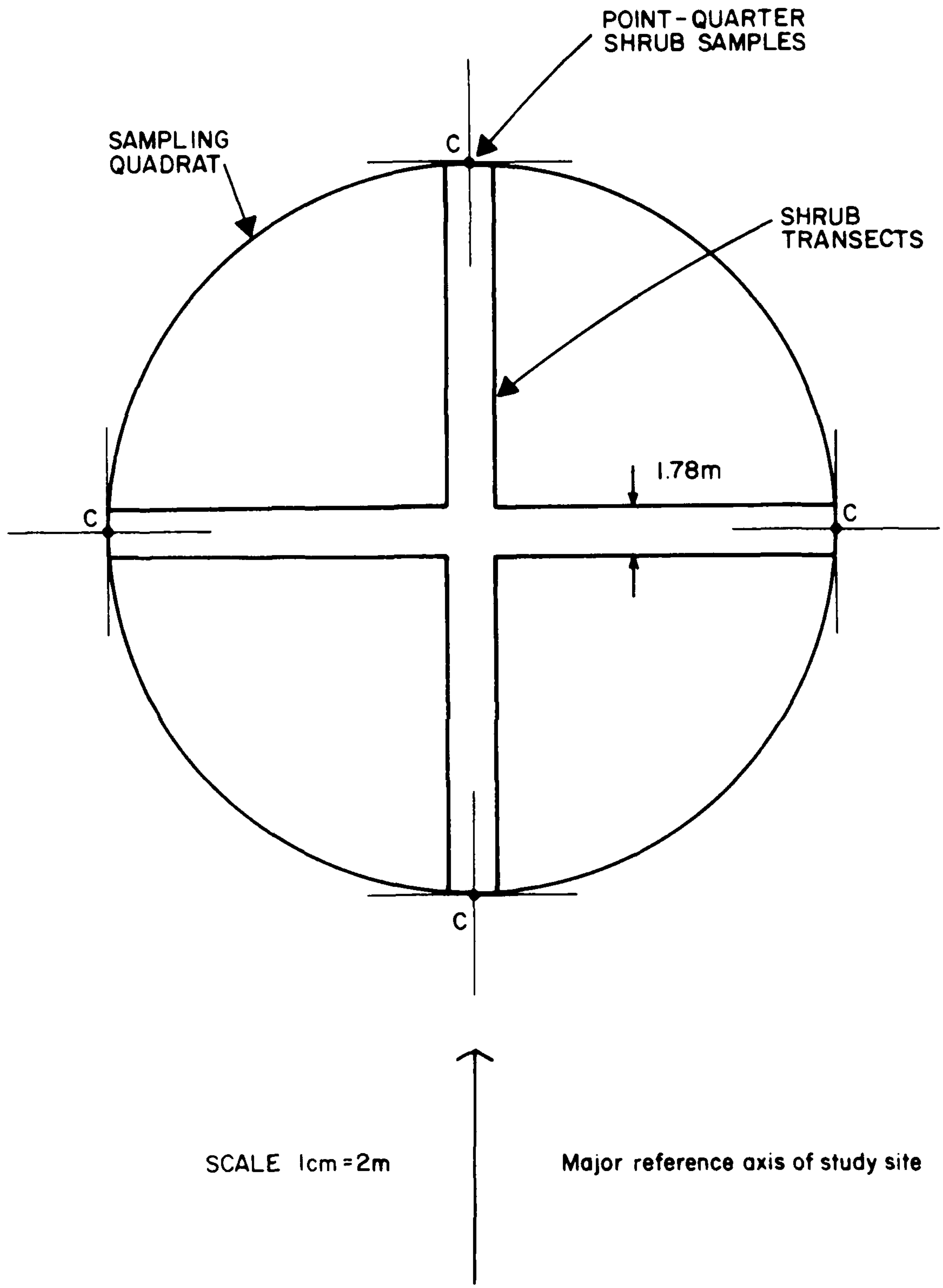


Figure 3. The 26-m-diameter sampling quadrat within each Illinois sampling unit, showing relative orientations of the shrub transects. The four point-quarter shrub sampling loci (c) are also shown.

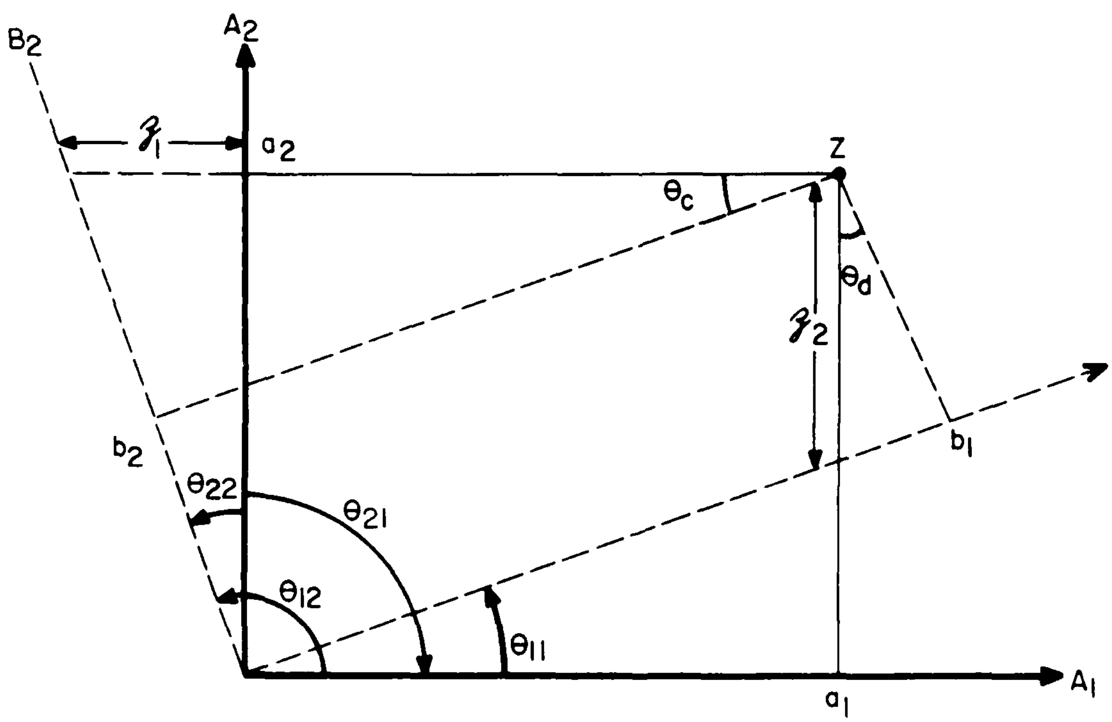


Figure 4. The rigid rotation of orthogonal axes; original axes A_1 and A_2 are rotated an angle of θ_{11} to form new reference axes B_1 and B_2.

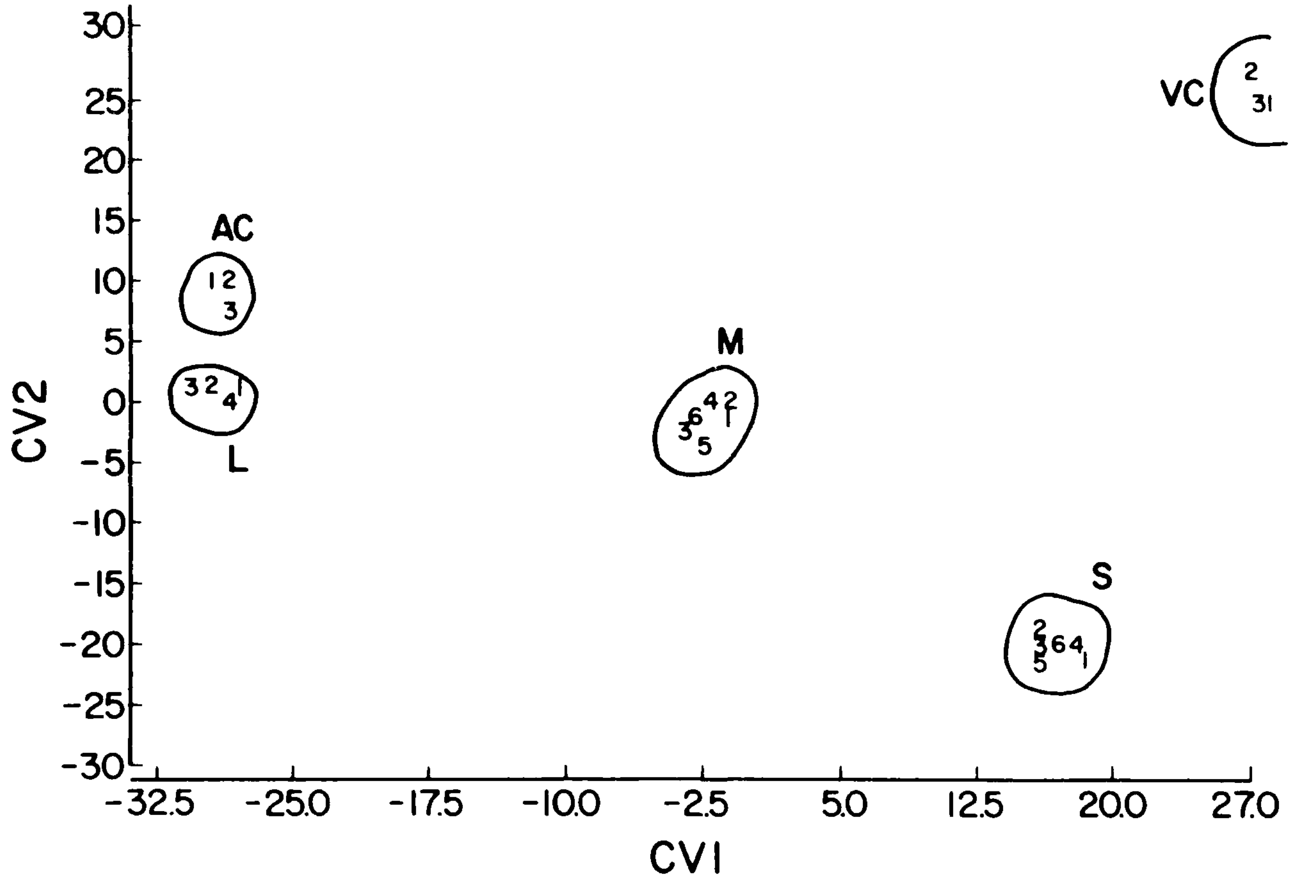


Figure 5. Ordination of Mojave Desert study sites in canonical space based on habitat variables passing the tolerance test. (See Table 1. Study sites are identified in Appenedix A. The numbers refer to individual transects.)

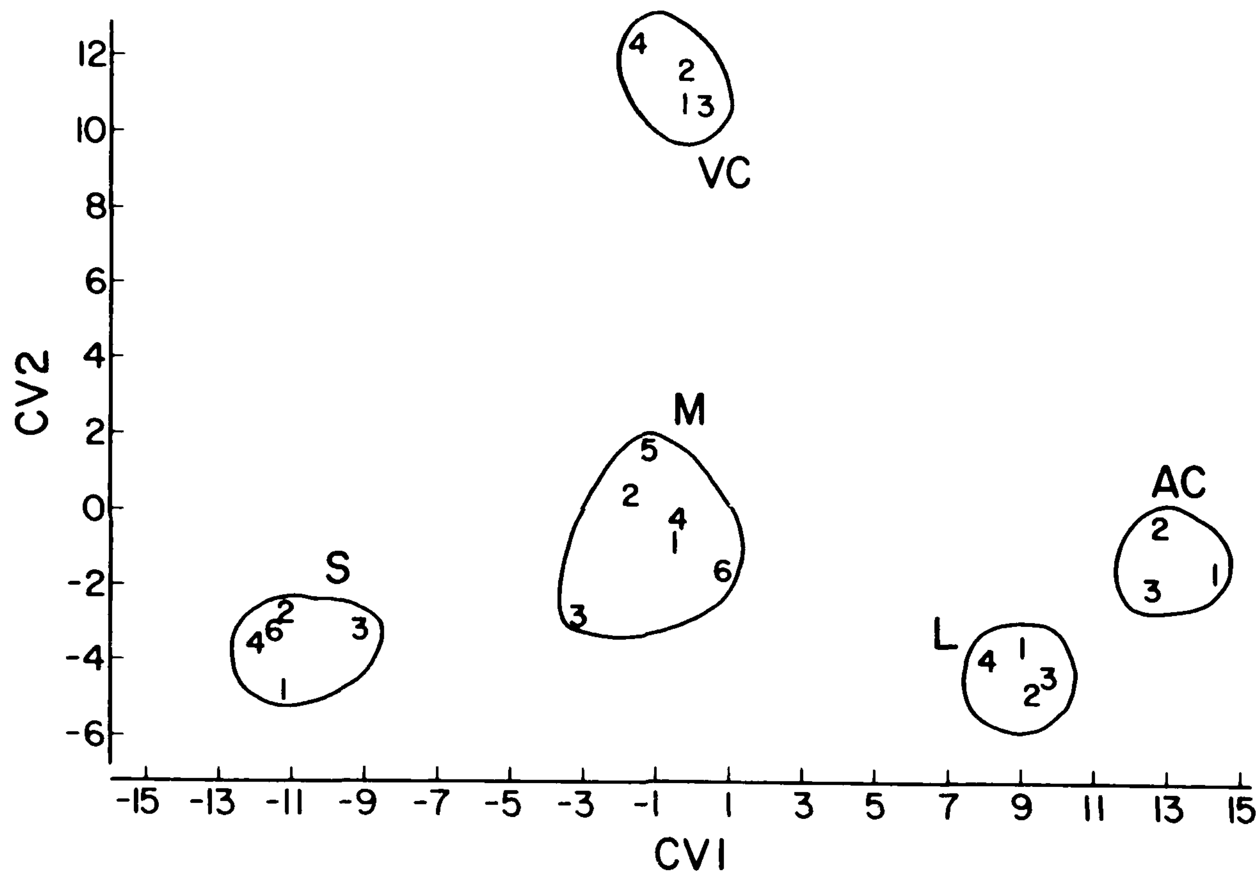


Figure 6. Ordination of Mojave Desert study sites in canonical space based on the "basic" habitat variables subset. (See Table 2. Study sites are identified in Appendix A. The numbers refer to individual transects.)

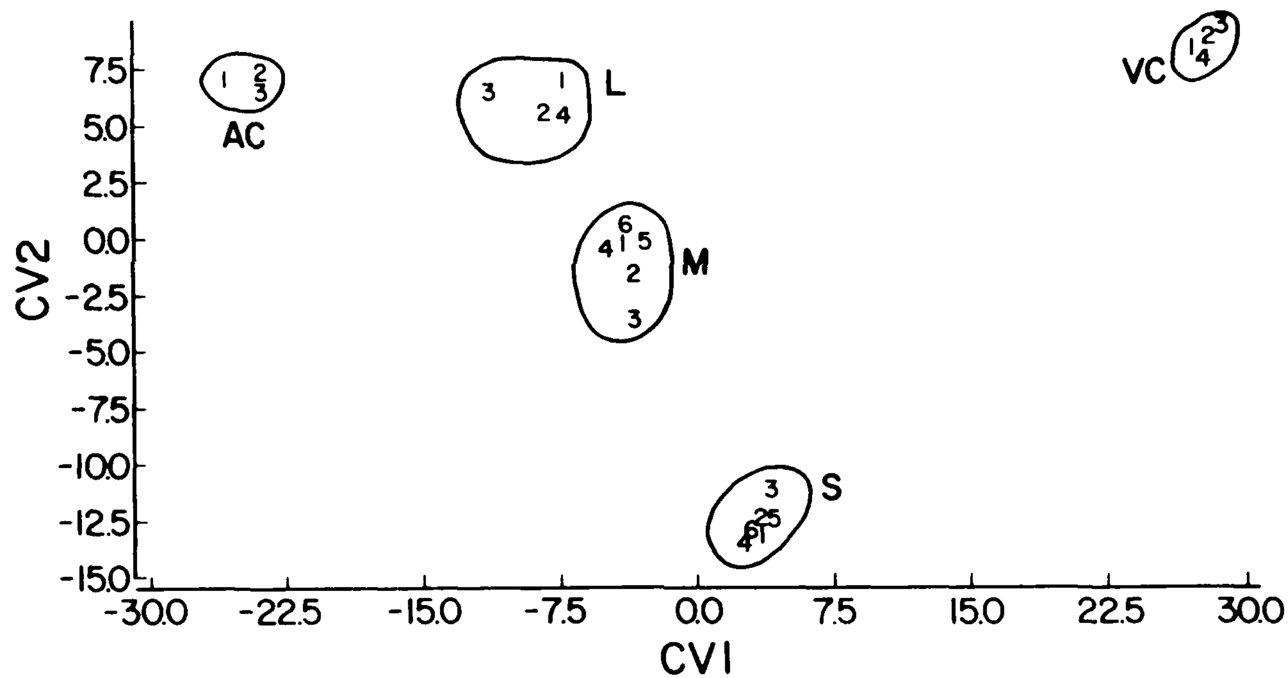


Figure 7. Ordination of Mojave Desert study sites based on the "basic" + shrub height variables subset. (See Table 3. Study sites are identified in Appendix A. The numbers refer to individual transects.)

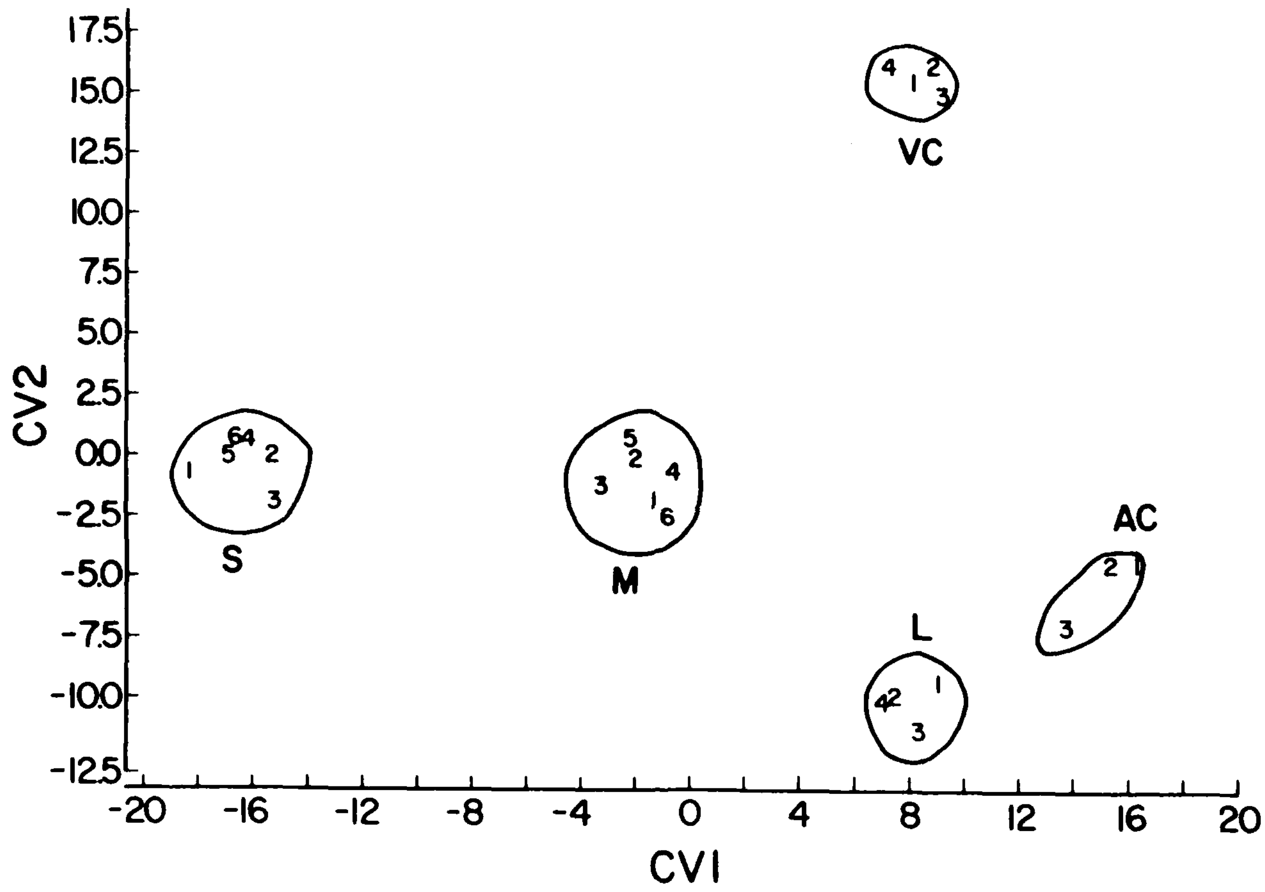


Figure 8. Ordination of Mojave Desert study sites based on the "basic" + shrub size heterogeneity variables subset. (See Table 4. Study sites are identified in Appendix A. The numbers refer to individual transects.)

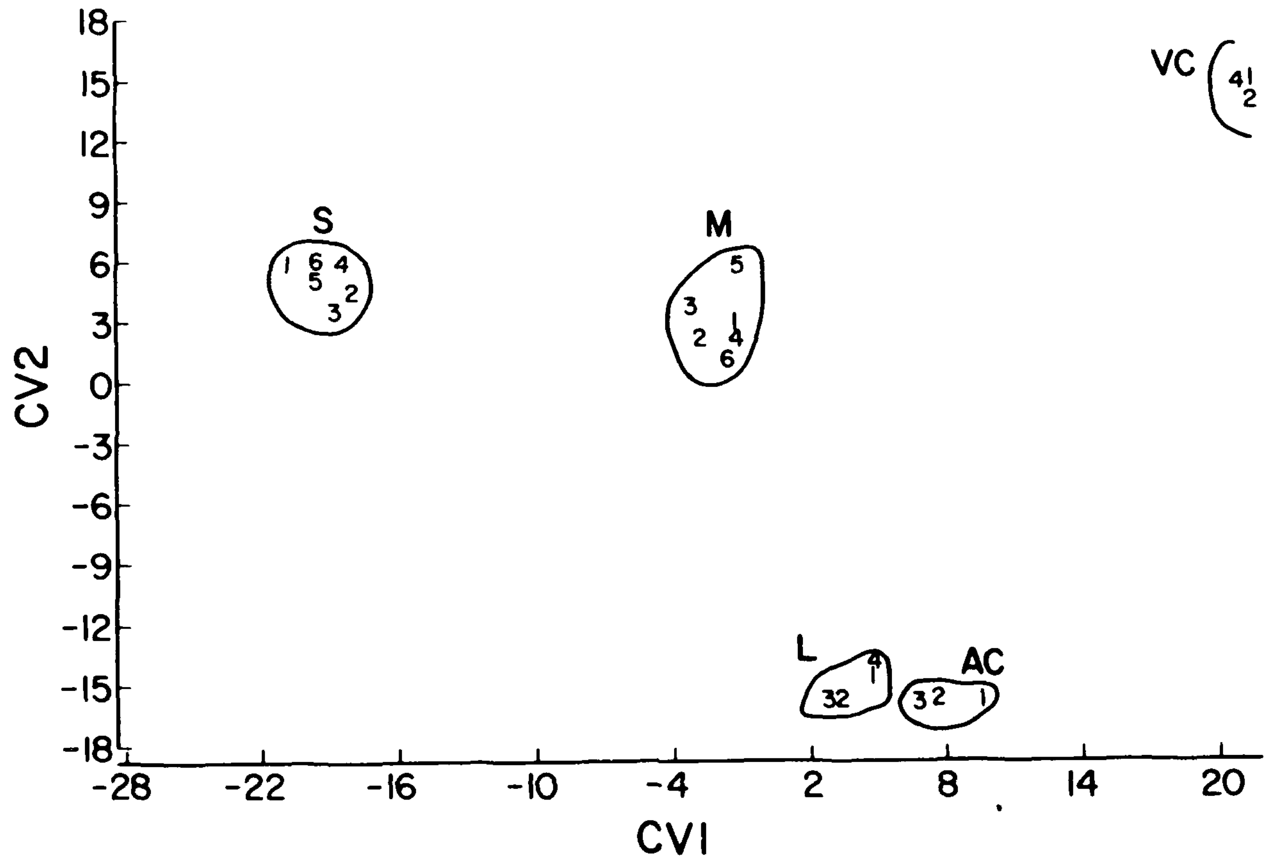


Figure 9. Ordination of Mojave Desert study sites based on the habitat variables from Figure 8 + mean height of dominant and subdominant shrub species + cover of shrubs $\geq$ 1.5 m in height. (See Table 6. Study sites are identified in Appendix A. The numbers refer to individual transects.)

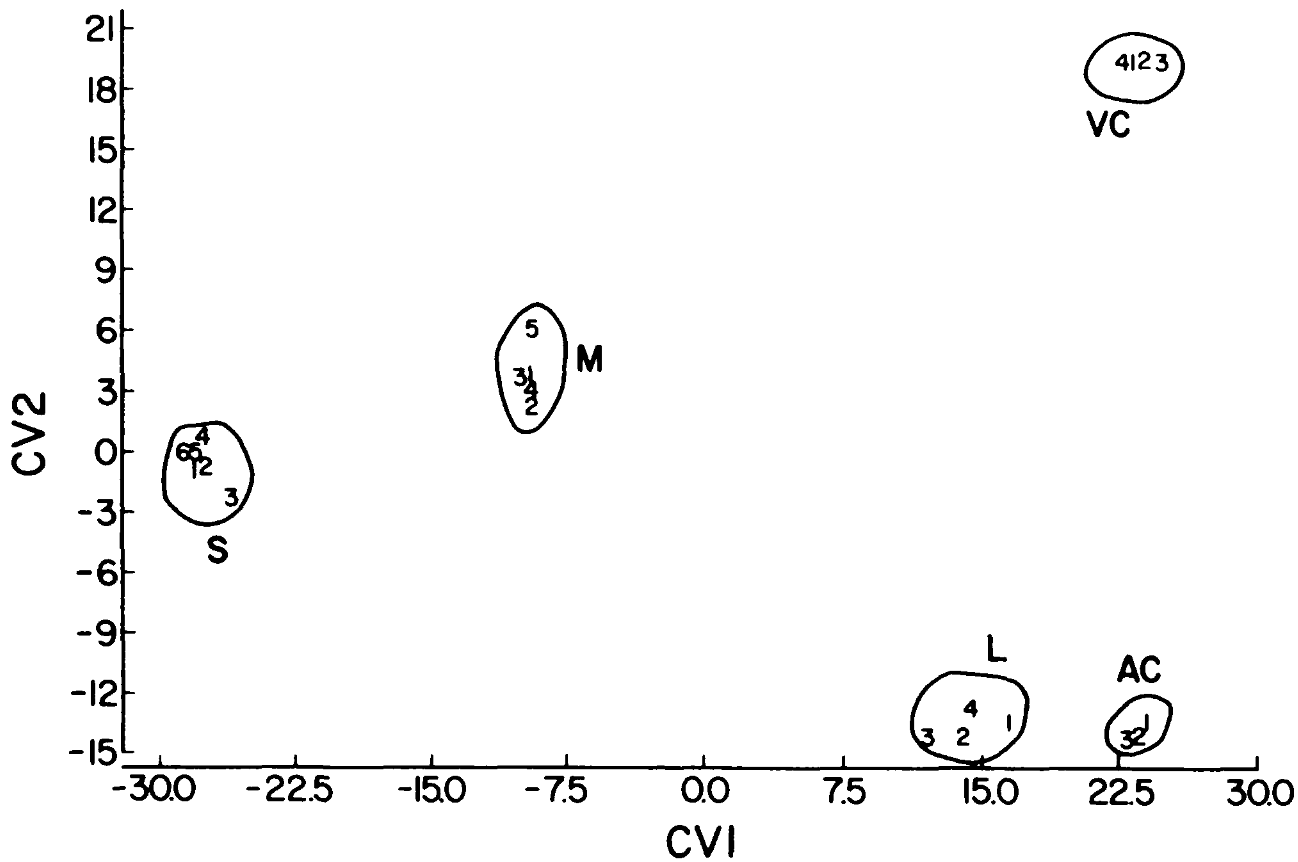


Figure 10. Ordination of Mojave Desert study sites based on the habitat variables from Figure 9 + density of shrubs 1.1 to < 1.5 m in height. (See Table 7. Study sites are identified in Appendix A. The numbers refer to individual transects.)

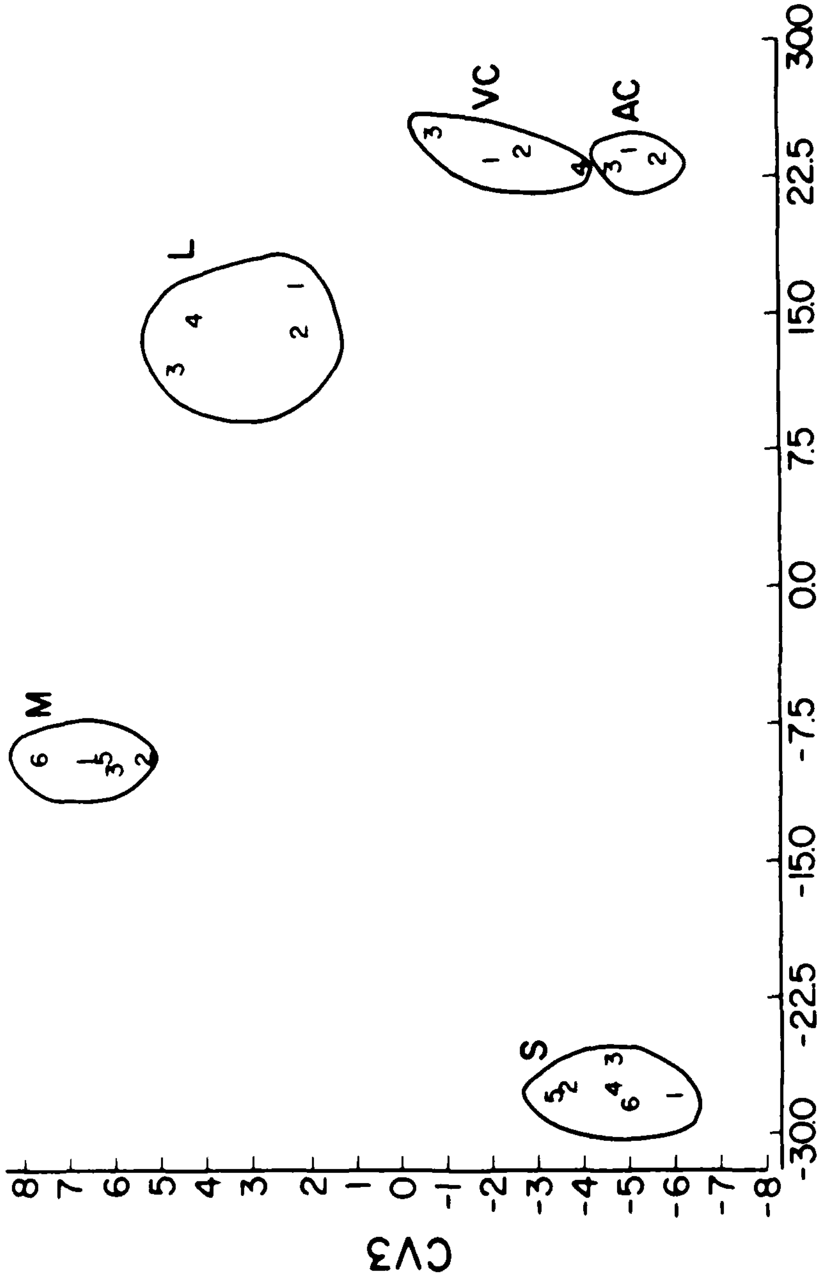


Figure 10. (Cont'd)

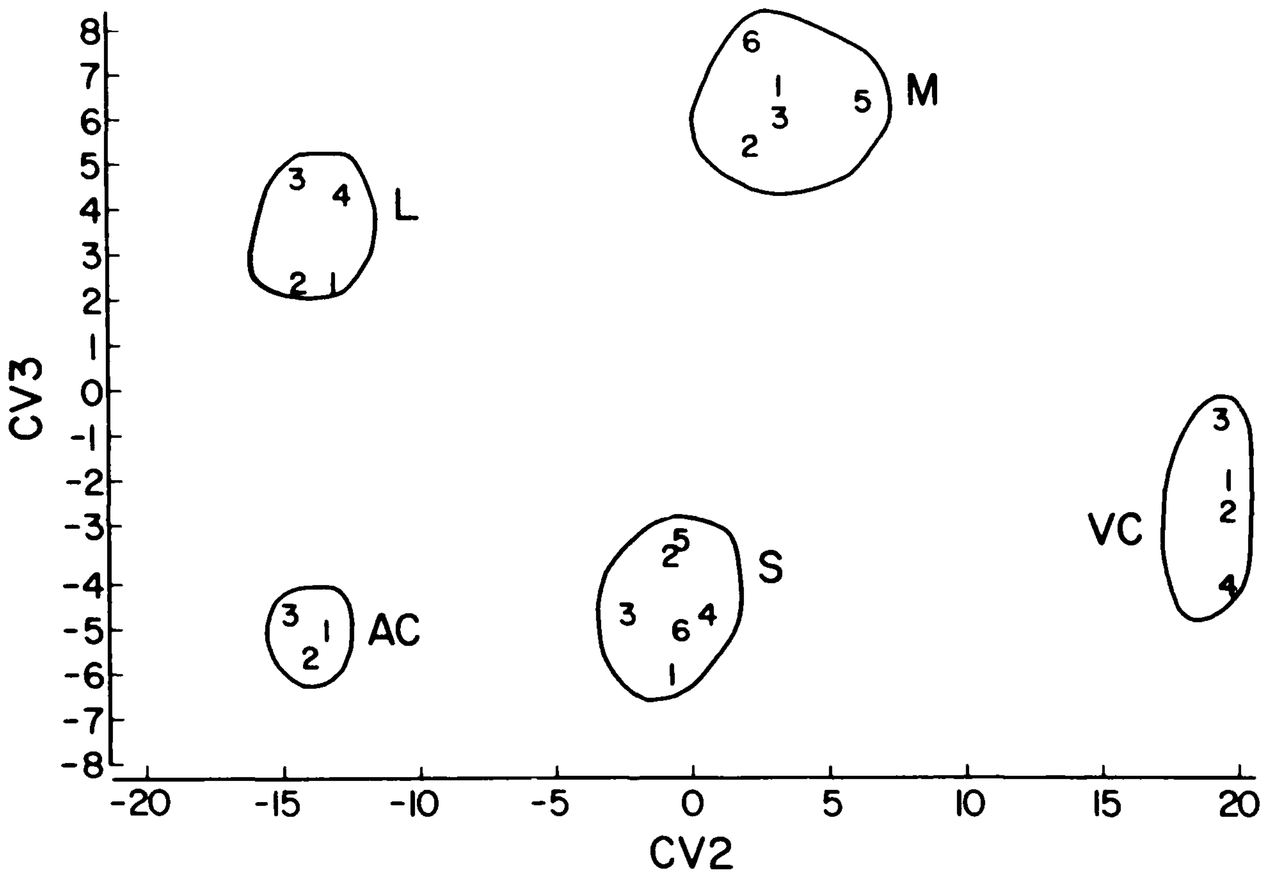


Figure 10. (Cont'd)

Table 1

Pearson Correlation of the First Four Canonical Variates With the Mojave Desert Habitat Variables That Passed the Tolerance Test

	CV 1	CV 2	CV 3	CV 4
CCB		0.84		
CBW		0.88		
CREST	-0.51[2]			
GRASS	0.58[2]			
FORB	-0.62[2]			
LIT				
GRAV	-0.92			
CSAND		0.79		
SAND	0.64[1]	-0.73		
ROCK				
CL4		0.77		
CVDCB		-0.89		
CVDBW	-0.76			
CVDR				
CVHCB		-0.89		
CVHBW	-0.58[2]		0.56[2]	
CVHR		0.53[2]		
CVNCB				

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 2

Pearson Correlation of the First Four Canonical Variates With the "Basic" Subset of Mojave Desert Habitat Variables

	CV 1	CV 2	CV 3	CV 4
CCB	0.78			
CBW	0.63[1]	0.64[1]		
CREST	0.56[2]			
GRASS		0.56[2]		
FORB	0.78			
LIT				0.51[2]
GRAV	0.91			
CSAND		0.88		
SAND	-0.94			
ROCK			-0.54[2]	

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 3

Pearson Correlation of the First Four Canonical Variates With the "Basic" + Shrub Height Subset of Mojave Desert Habitat Variables

	CV 1	CV 2	CV 3	CV 4
CCB		0.91		
CBW		0.86		
CREST				
GRASS	0.51[2]			
FORB		0.59[2]		
LIT		0.52[2]		
GRAV	-0.76	0.57[2]		
CSAND	0.69[1]	0.59[2]		
SAND		-0.92		
ROCK				-0.51[2]
L2	-0.52[2]		0.68[1]	
L4J		0.73		
L8		0.55[2]		
L10		0.73		
L12J		0.92		
CL4		0.83		

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 4

Pearson Correlation of the First Four Canonical Variates With the "Basic" + Shrub Size Heterogeneity Subset of Mojave Desert Habitat Variables

	CV 1	CV 2	CV 3	CV 4
CCB	0.91			
CBW	0.82			
CREST	0.51[2]			
GRASS		0.64[1]		
FORB	0.70[1]			
LIT				
GRAV	0.71	-0.65[1]		
CSAND		0.74		
SAND	-0.95			
ROCK				
CVDCB	-0.85			
CVDBW		-0.78		

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 5

Pearson Correlation of the First Four Canonical Variates With the Habitat Variables From Table 4 + Cover of Shrubs $\geq$ 1.5 m in Height

	CV 1	CV 2	CV 3	CV 4
CCB	0.90			
CBW	0.81			
CREST	0.51[2]			
GRASS	0.64[1]	0.64[1]		
FORB	0.72			
LIT				
GRAV	0.73	-0.63[1]		
CSAND		0.76		
SAND	-0.95			
ROCK				
CVDCB	-0.84			
CVDBW		-0.77		
CL4	0.77			

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 6

Pearson Correlation of the First Four Canonical Variates With the Habitat Variables From Table 5 + Mean Heights of Dominant and Subdominant Shrub Species

	CV 1	CV 2	CV 3	CV 4
CCB	0.87			
CBW	0.88			
CREST				
GRASS		-0.62[2]		
FORB		0.68[1]		
LIT	0.52[2]			
GRAV		0.92		
CSAND	0.76			
SAND	-0.78	-0.57[2]		
ROCK			0.53[2]	
CVDCB	-0.91			
CVDBW		0.72		
CL4	0.80			
HCB	0.92			
HBW	0.83			
HREST				

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 7

Pearson Correlation of the First Four Canonical Variates With the Habitat Variables From Table 6 + Density of Shrubs 1.1 to < 1.5 m in Height

	CV 1	CV 2	CV 3	CV 4
CCB	0.94			
CBW	0.83			
CREST				
GRASS		0.65[1]		
FORB	0.68[1]			
LIT	0.52[2]			
GRAV	0.58[2]	-0.76		
CSAND	0.58[2]	0.68[1]		
SAND	-0.90			
ROCK			0.59[2]	
CVDCB	-0.89			
CVDBW		-0.71		
L12J	0.93			
CL4	0.78			
HCB	0.89			
HBW	0.80			
HREST				

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

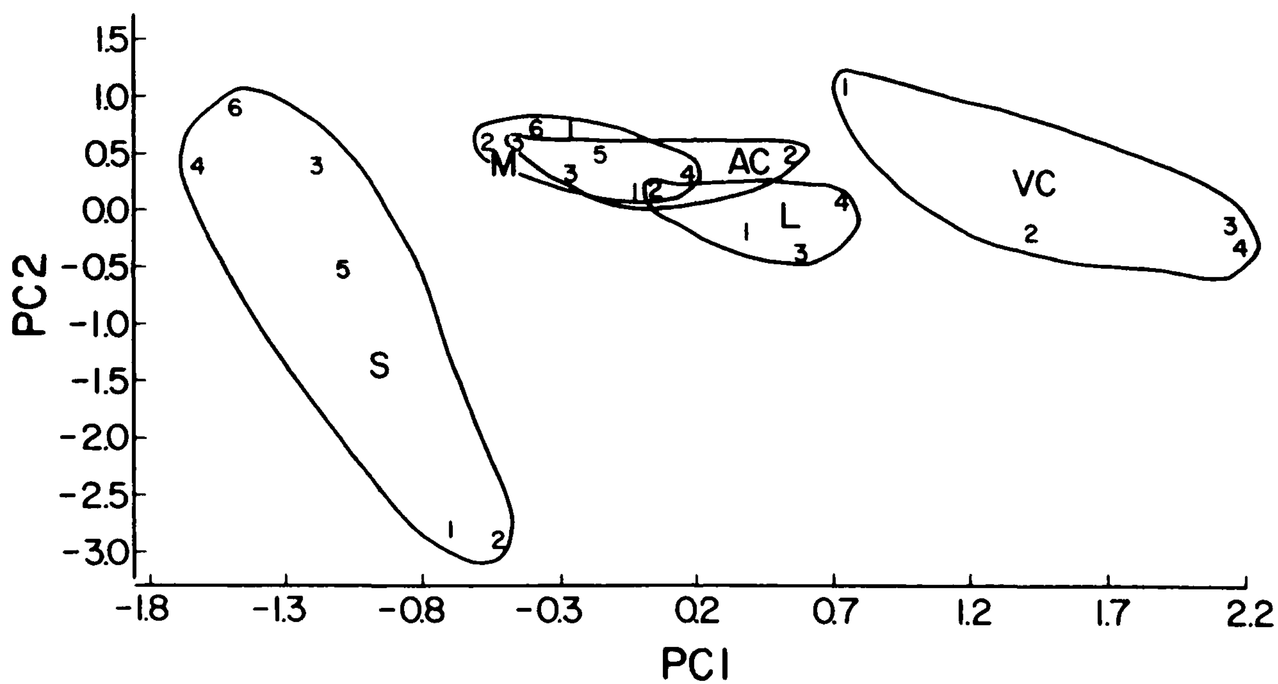


Figure 11. Ordination of Mojave Desert study sites in principal components space based on the habitat variables that formed the optimal subset for CAD. (See Table 8. Study sites are identified in Appendix A. The numbers refer to individual transects.)

Table 8

Pearson Correlation of the First Four Principal Components With the Mojave Desert Habitat Variables That Formed the Optimal Subset for CAD (Table 7 Variables)

	PC 1	PC 2	PC 3	PC 4
CCB	0.78			
CBW	0.81			
CREST		0.82		
GRASS		0.55[2]		-0.58[2]
FORB			0.90	
LIT	0.70[1]			
GRAV			0.86	
CSAND	0.88			
SAND	-0.65[1]		-0.56[2]	
ROCK				0.87
CVDCB	-0.89			
CVDBW				0.74
L12J	0.74			
CL4	0.69[1]	0.64[1]		
HCB	0.82			
HBW	0.89			
HREST		0.91		

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

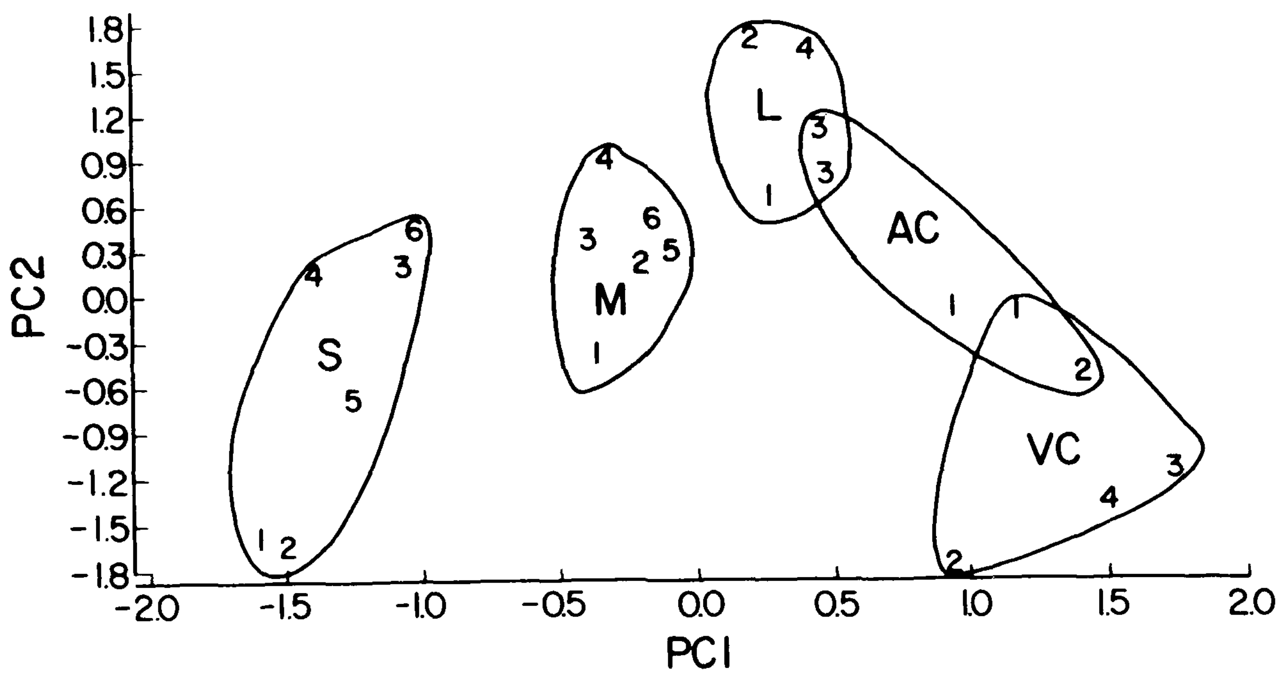


Figure 12. Ordination of Mojave Desert study sites based on the complete set of habitat variables. (See Table 9. Study sites are identified in Appendix A. The numbers refer to individual transects.)

Table 9

Pearson Correlation of the First Four Principal Components With the Complete Set of Mojave Desert Habitat Variables

	PC 1	PC 2	PC 3	PC 4
CCB	0.94			
CBW	0.89			
CCHE		0.75		
CSH	0.75			
CTB		0.59[2]		
CMT			0.69[1]	
CCAS			0.80	
CMD			0.71	
CA	0.70[1]			
CCD		0.62[2]		
C123				0.62[2]
CREST		0.70[1]		
CL1	0.81			
CL2	0.92			
CL3	0.89			
CL4	0.84			
HCB	0.90			
HBW	0.83			
HCHE		0.73		
HSH	0.77			
HTB		0.52[2]		
HMT			0.57[2]	
HCAS			0.77	
HMD			0.81	
HA	0.67[1]			
HCD		0.59[2]		
H123				0.63[1]
HREST	0.55[2]			
GRASS				0.68[1]
FORB	0.52[2]			0.57[2]
LIT	0.62[2]			
GRAV		0.57[2]		-0.62[2]
CSAND	0.70[1]			
SAND	-0.81			
ROCK		0.63[1]		
CVDCB	-0.88			
CVDBW		0.61[2]		
CVDR				
CVHCB	-0.90			
CVHBW		0.70[1]		
CVHR				
CVNCB				
CVNBW			-0.54[2]	
CVNR				
L2			0.74	
L4	0.74			
L6	0.88			
L8	0.53[2]	0.65[1]		
L10	0.72			
L12	0.87			
L14	0.93			
L16	0.74			
L18	0.69[1]			
L20				
L22				
L27				
SHCOV	0.98			
NSP		0.89		

P $\leq$ 0.0001

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

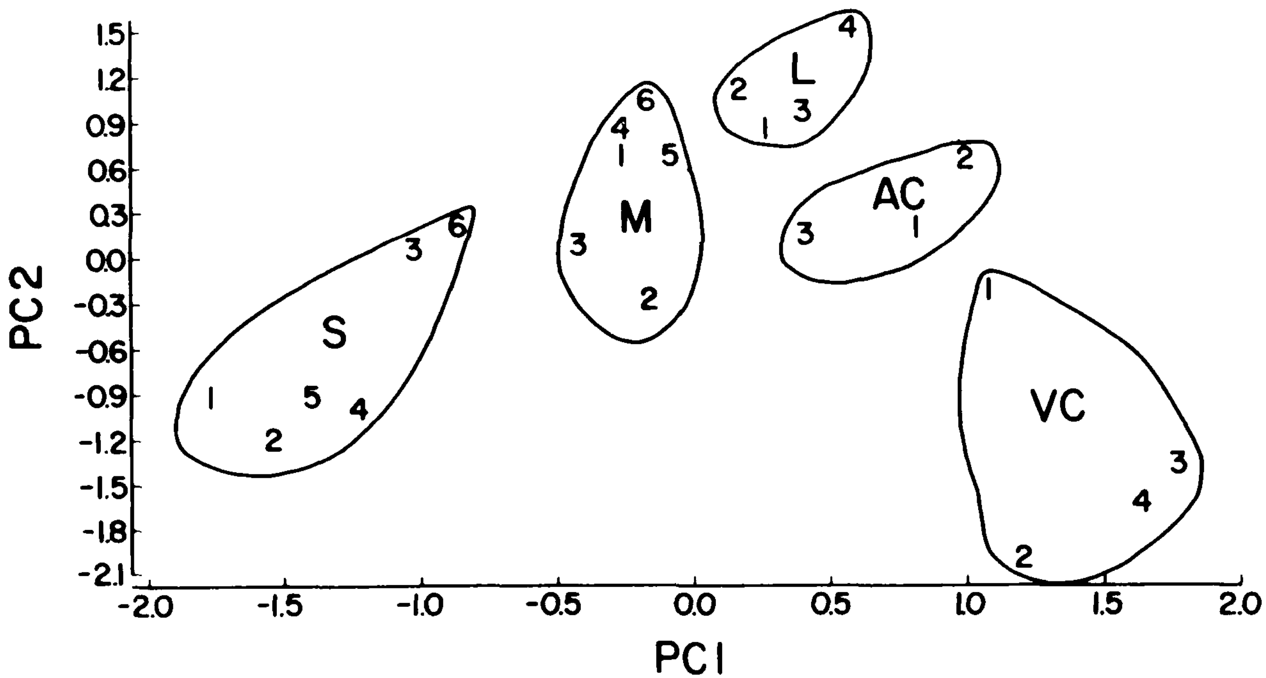


Figure 13. Ordination of Mojave Desert study sites based on the habitat variables that formed the optimal subset for PCA. (See Table 10. Study sites are identified in Appendix A. The numbers refer to individual transects.)

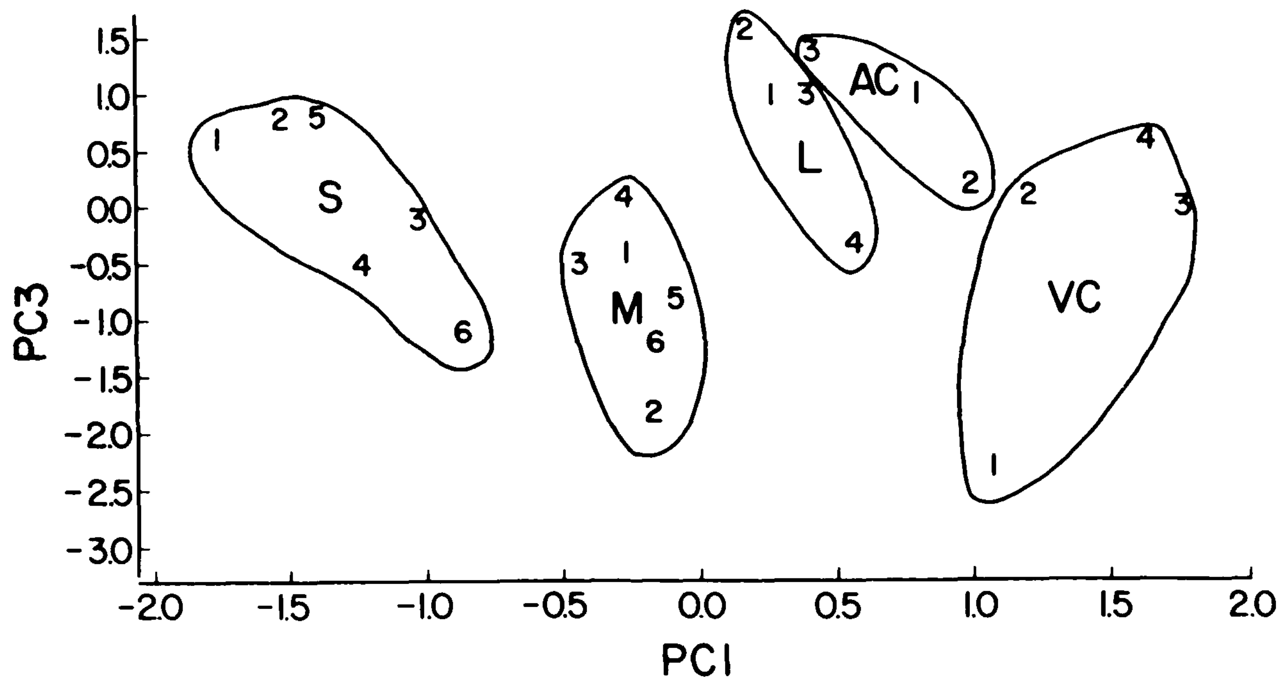


Figure 13. (Cont'd)

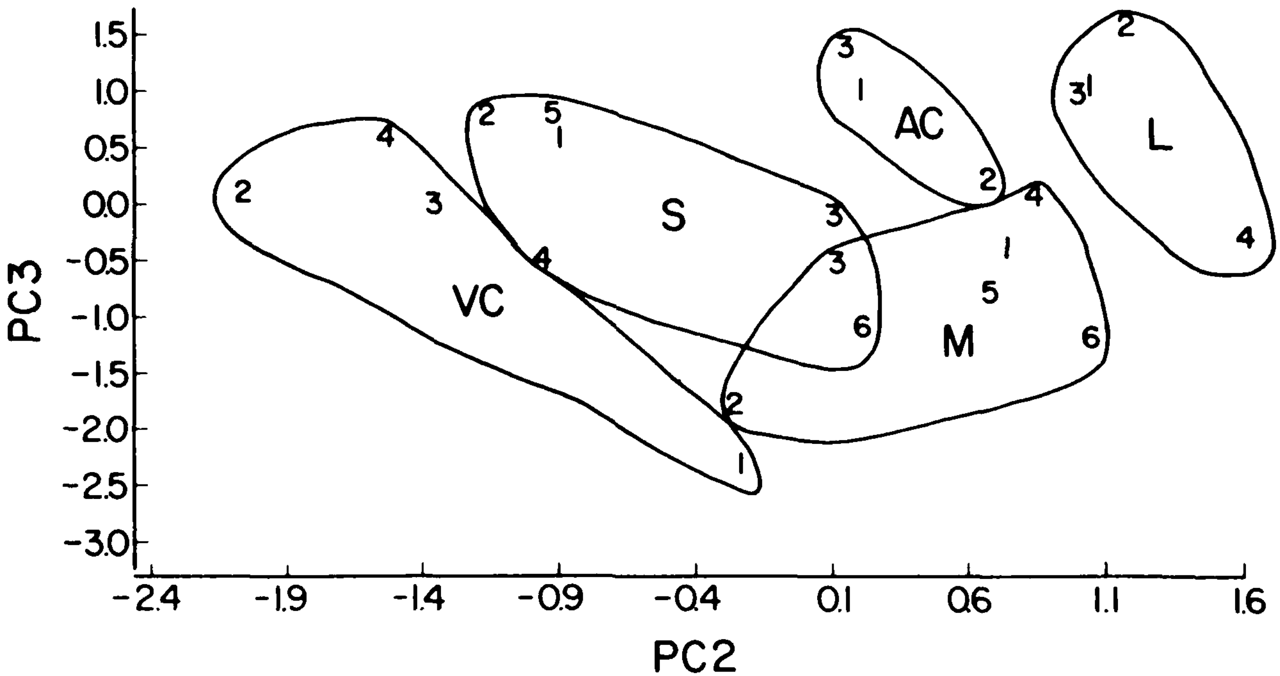


Figure 13. (Cont'd)

Table 10

Pearson Correlation of the First Four Principal Components With the Mojave Desert Habitat Variables That Formed the Optimal Subset for PCA

	PC 1	PC 2	PC 3	PC 4
CCB	0.93			
CBW	0.87			
CREST		0.64[1]		
GRASS			-0.75	
FORB			0.72	
LIT	0.62[2]			
GRAV		0.62[2]		
CSAND	0.75			
SAND	-0.80			
ROCK		0.60[2]		
CVDCB	-0.88			
CVDBW		0.86		
CVDR				
CVHCB	-0.90			
CVHBW		0.85		
CVHR				
CVNCB			-0.55[2]	
CVNBW			0.58[2]	-0.59[2]
CVNR			0.61[2]	
L2				0.81
L4J	0.82			
L8		0.57[2]		
L10	0.68[1]		0.56[2]	
L12J	0.92			
CL4	0.85			

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

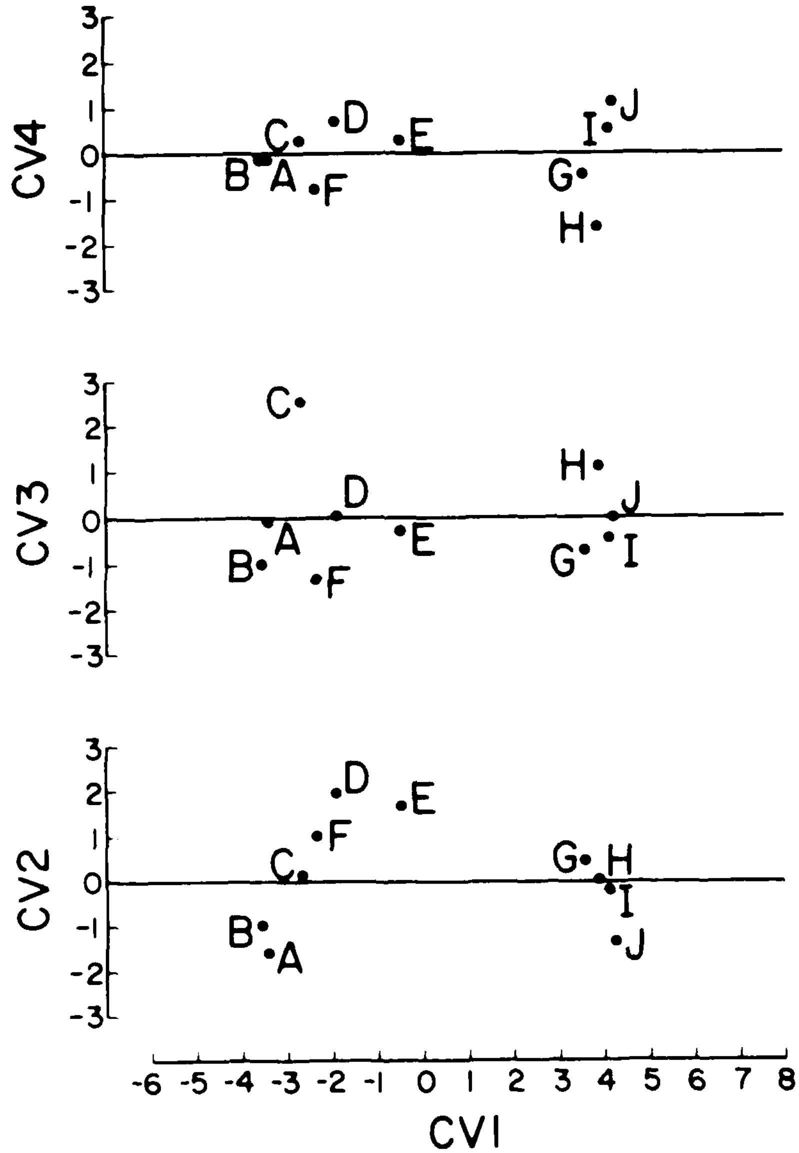


Figure 14. Ordination of Illinois study sites in canonical space based on the structural subset of habitat variables. (See Table 11. Study sites are identified in Appendix D. The loci represent mean canonical variate scores.)

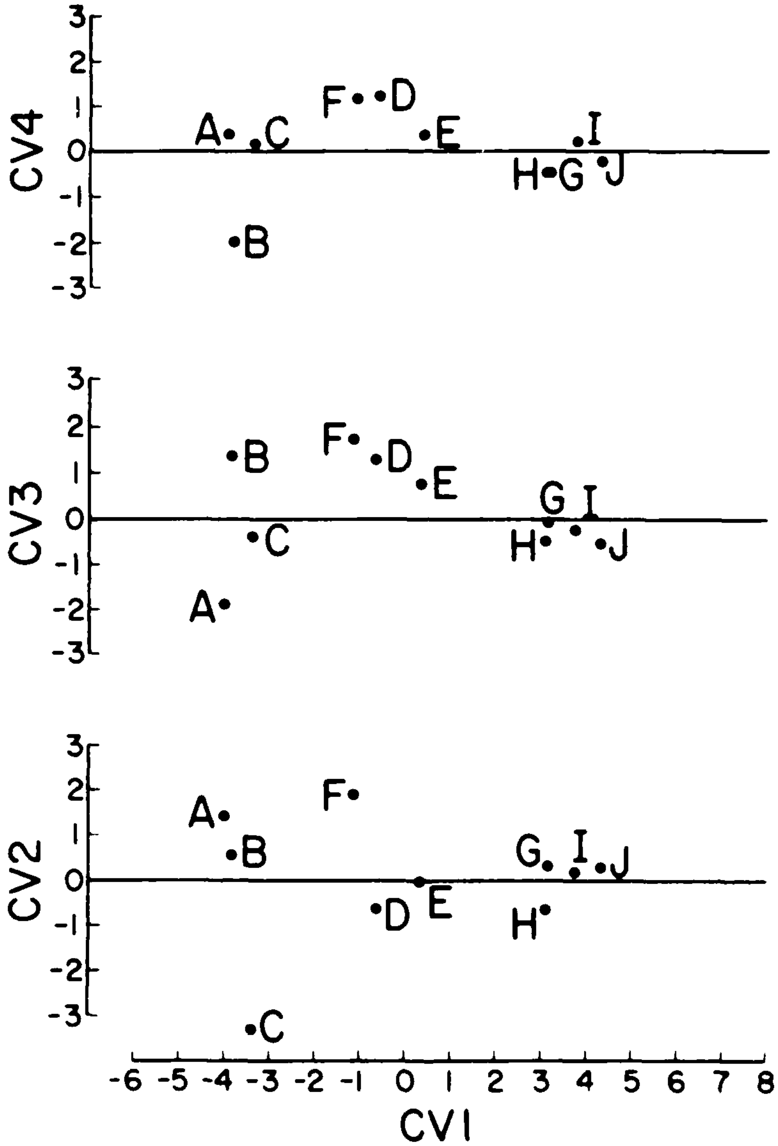


Figure 15. Ordination of Illinois study sites in canonical space based on the typical subset of habitat variables. (See Table 12. Study sites are identified in Appendix D. The loci represent mean canonical variate scores.)

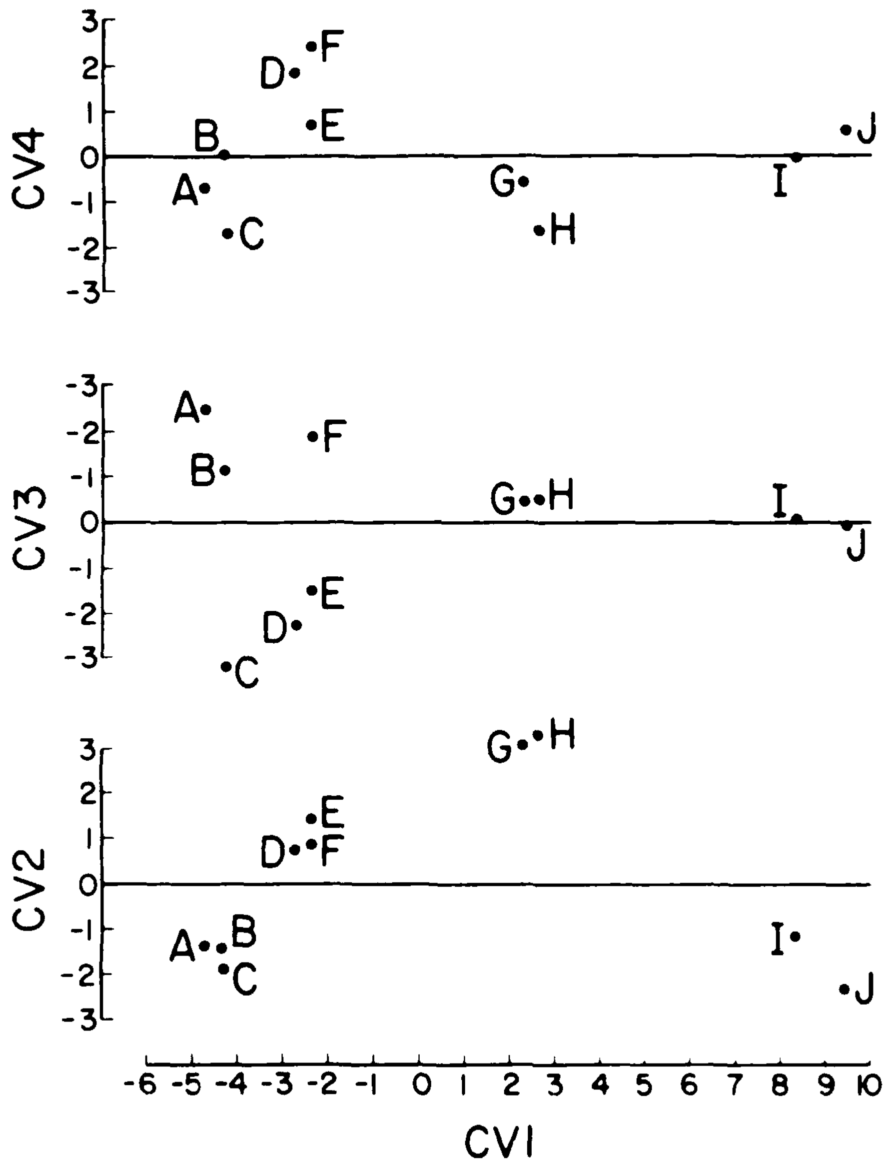


Figure 16. Ordination of Illinois study sites in canonical space based on the typical + tree guilds subset of habitat variables. (See Table 13. Study sites are identified in Appendix D. The loci represent mean canonical variate scores.)

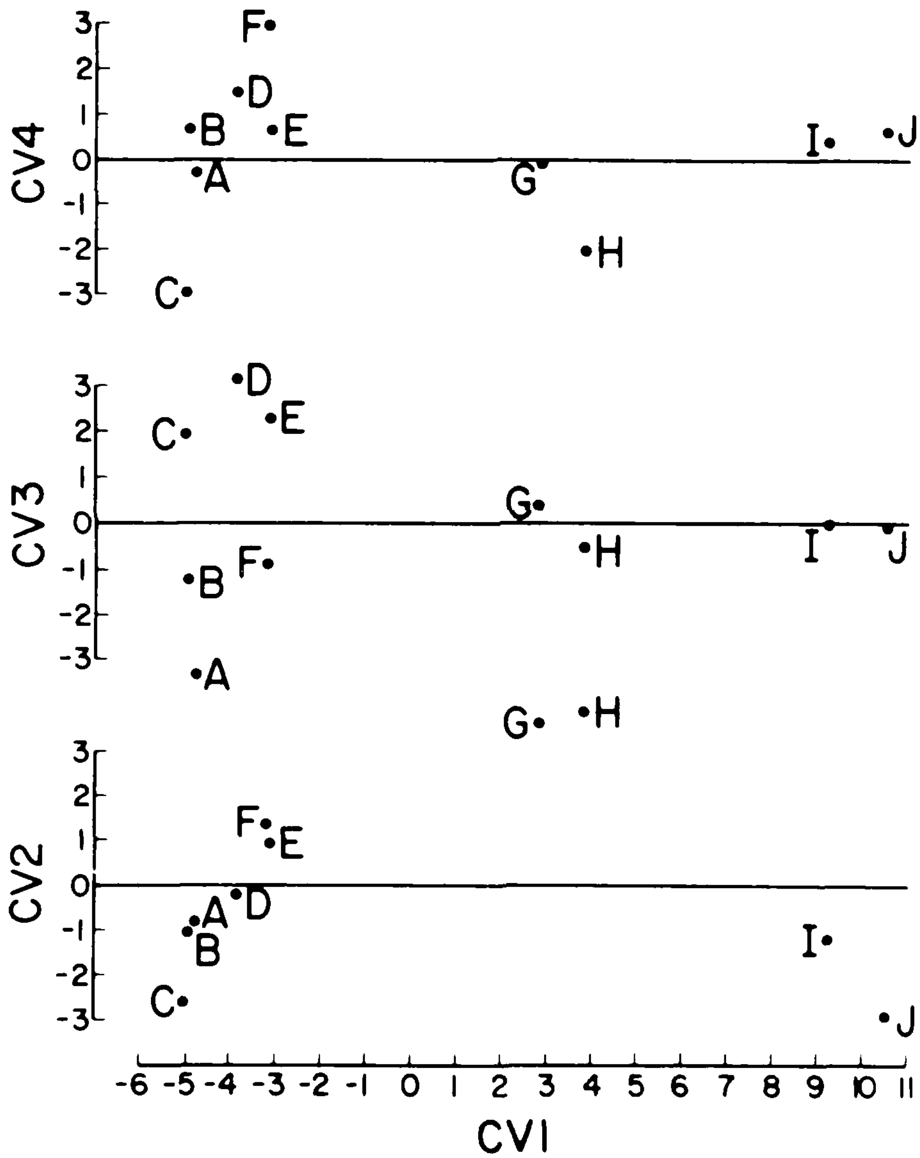


Figure 17. Ordination of Illinois study sites in canonical space based on the complete set of habitat variables. (See Table 14. Study sites are identified in Appendix D. The loci represent mean canonical variate scores.)

Table 11

Pearson Correlation of the First Four Canonical Variates With the Structural Subset of Illinois Habitat Variables

	CV 1	CV 2	CV 3	CV 4
L1	-0.16[2]			
L2	-0.23		-0.47	0.46
L3	0.27	0.22		0.35
L4	0.38	0.41	0.28	0.39
L5	0.36	0.37	0.48	0.26
L6	0.48	0.50	0.38	0.19[1]
L7	0.62	0.49		0.14[2]
L8	0.71	0.50	-0.21	
L9	0.81	0.47	-0.18[1]	
L10	0.91	0.21		
L11	0.98	-0.14[2]		
VEGVOL	0.63	0.29		
VHW	0.76	0.50	0.14[2]	0.19[1]
HZHW		0.38		0.24
GC	-0.30			
CC	0.71	0.49	0.30	0.20
AQUA		0.13[2]		-0.58
TOP		-0.39	0.37	0.14[2]

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 12

Perarson Correlation of the First Four Canonical Variates With the Typical Subset of Illinois Habitat Variables

	CV1	CV2	CV3	CV4
CC	-0.29	0.20	-0.18[1]	0.27
CC	0.75	-0.43		0.29
SHRD	0.13[2]		0.34	0.49
SHRH	-0.13[2]		0.32	0.35
SHRC	0.20		0.15[2]	0.51
NSPSHR	0.46	0.50		0.48
CANAV	0.92	-0.27	-0.18[1]	
CANMX	0.93	-0.16[2]		
NSPTRE	0.81	0.13[2]		
SAP	0.35	-0.50	-0.17[1]	0.61
A	0.58	-0.32	0.28	0.55
B	0.83		0.29	0.18[1]
C	0.85			
D	0.79			-0.17[1]
E	0.76		-0.24	-0.27
F	0.47		-0.14[2]	-0.23
G	0.33			-0.16[2]
H	0.20[1]			
I				
TREMAS	0.92	-0.23		0.15[2]

$P < 0.0001$

1 $0.0001 < P < 0.001$

2 $0.001 < P \leq 0.01$

Table 13

Pearson Correlation of the First Four Canonical Variates With the Typical + Tree Guilds Subsets of Illinois Habitat Variables

	CV 1	CV 2	CV 3	CV 4
GC	-0.30		0.18[1]	
CC	0.60	0.39	-0.50	
SHRD	0.15[2]		-0.18[1]	0.56
SHRH	-0.14[2]		-0.14[2]	0.40
SHRC	0.22		-0.13[2]	0.43
NSPSHR	0.35	0.33	0.25	0.36
CANAV	0.81	0.35	-0.30	-0.16[2]
CANMX	0.78	0.46	-0.25	
NSPTRE	0.71	0.35		0.22
SAP	0.30		-0.49	
A	0.41	0.42	-0.48	0.24
B	0.67	0.50	-0.14[2]	0.30
C	0.76	0.34		0.16[2]
D	0.75	0.21		
E	0.73	0.22		-0.18[1]
F	0.37	0.41		-0.28
G	0.26	0.34		-0.22
H	0.18[1]			-0.15[2]
I				
TREMAS	0.78	0.43	-0.29	
DUP	0.86	-0.42		0.14[2]
DBOT	0.27	0.72		-0.29
DRICH	0.72	-0.13[2]		
DEAR	0.13	0.31	0.34	0.14[2]
DDEA	0.44	0.35		0.35
DBL	-0.35		-0.85	
DC		0.69	-0.17[1]	
DELM	0.45	0.57	0.17[1]	-0.27

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

Table 14

Pearson Correlation of the First Four Canonical Variates With the Complete Set of Illinois Habitat Variables

	CV 1	CV 2	CV 3	CV 4
L1	-0.16[2]			
L2	-0.13[2]	-0.22		0.51
L3	0.28		0.27	
L4	0.32		0.49	
L5	0.28		0.49	-0.25
L6	0.36	0.14[2]	0.57	-0.18[1]
L7	0.48	0.30	0.48	
L8	0.57	0.44	0.39	0.17[1]
L9	0.66	0.48	0.37	0.13[2]
L10	0.77	0.44	0.24	
L11	0.91	0.28		-0.14[2]
VEGVOL	0.54	0.25	0.29	
VHW	0.62	0.29	0.52	
HZHW			0.36	0.13[2]
GC	-0.29		-0.19[1]	
CC	0.58	0.23	0.57	-0.15[2]
AQUA		0.33		
TOP		-0.21	-0.17[1]	-0.35
SHRD		-0.19[1]	0.27	0.44
SHRH	-0.17[1]		0.19[1]	0.31
SHRC	0.19[1]	-0.17[1]	0.20	0.34
NSPSHR	0.34	0.32		0.41
CANAV	0.81	0.24	0.37	-0.19[1]
CANMX	0.77	0.35	0.39	
NSPTRE	0.70	0.28	0.19[1]	0.20
SAP	0.28		0.43	-0.25
A	0.38	0.26	0.60	
B	0.65	0.40	0.36	0.27
C	0.74	0.26	0.25	0.16[2]
D	0.75	0.18[1]		
E	0.74	0.21		
F	0.40	0.42		-0.20
G	0.28	0.35		-0.16[2]
H	0.19[1]	0.13[2]		
I				
TREMAS	0.77	0.32	0.41	
DUP	0.84	-0.44		0.18[1]
DBOT	0.29	0.69		-0.24
DRICH	0.71			0.13[2]
DEAR	0.14[2]	0.35	-0.21	0.19[1]
DDEA	0.42	0.27	0.28	0.26
DBL	-0.38	-0.26	0.72	-0.31
DC		0.62	0.32	
DELM	0.47	0.59		-0.16[2]

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

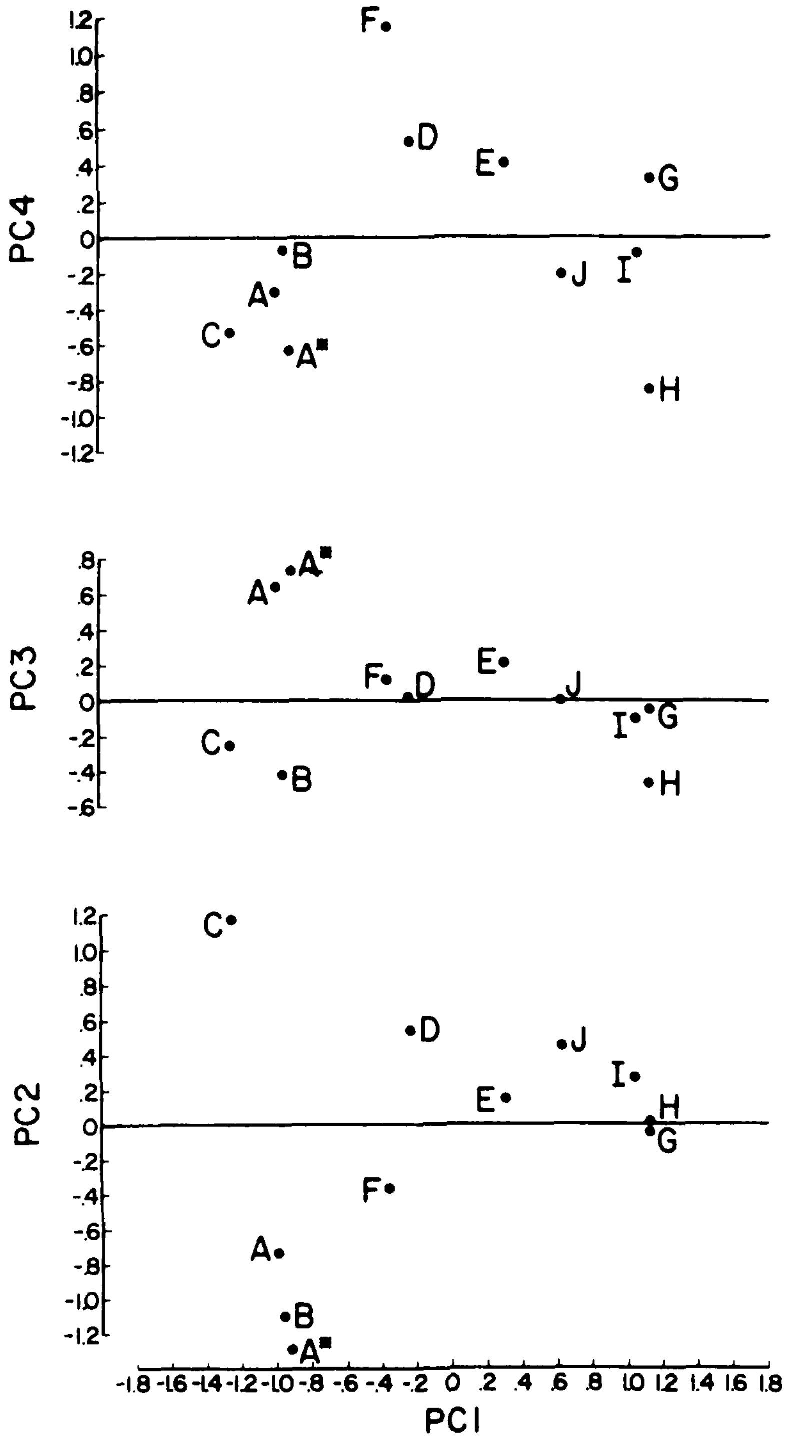


Figure 18. Ordination of Illinois study sites in principal components space based on the structural variable subset of habitat variables. (See Table 15. Study sites are identified in Appendix D. The loci represent mean principal components scores. A* represents study site A after the sapling edge was deleted from the analysis.)

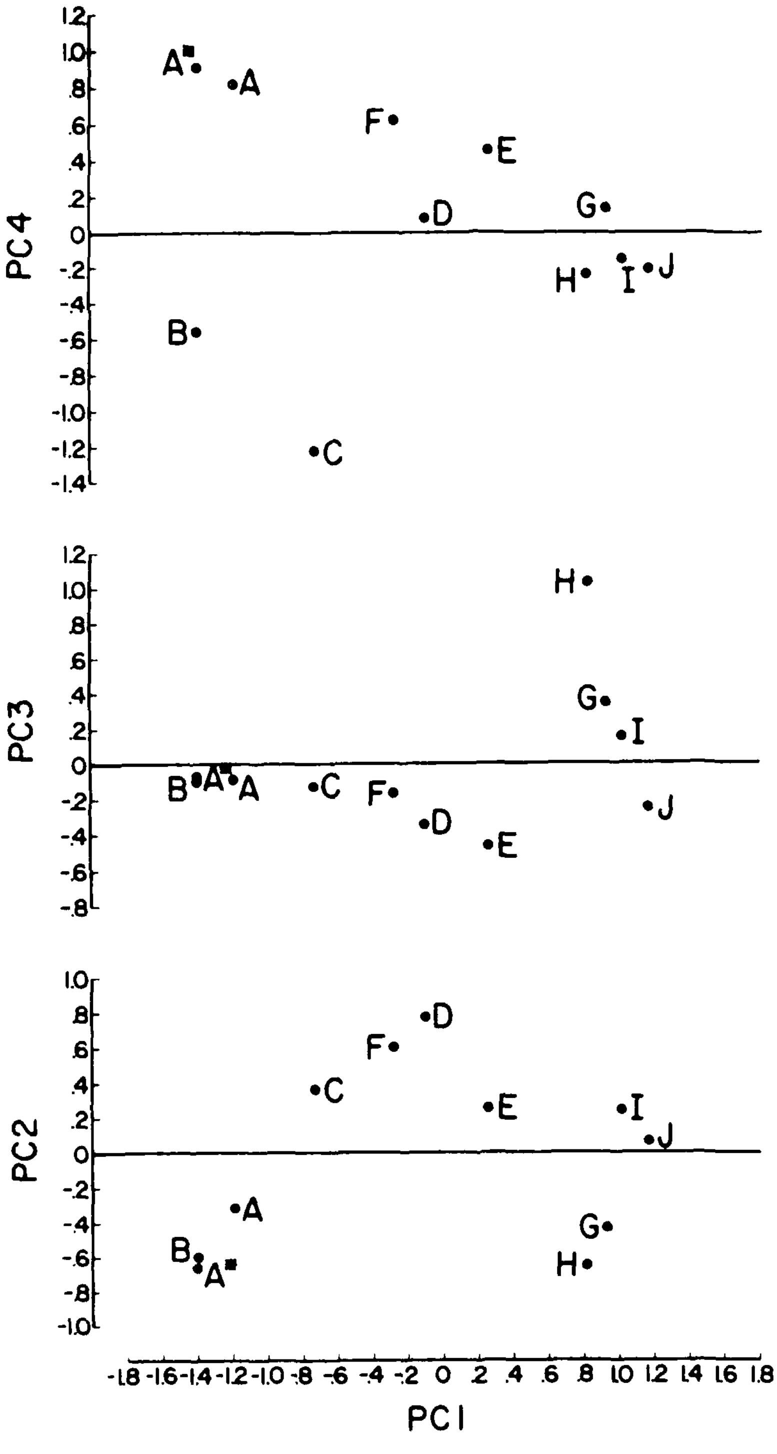


Figure 19. Ordination of Illinois study sites in principal components space based on the typical variable subset of habitat variables. (See Table 16. Study sites are identified in Appendix D. The loci represent mean principal components scores. A* represents study site A after the sapling edge was deleted from the analysis.)

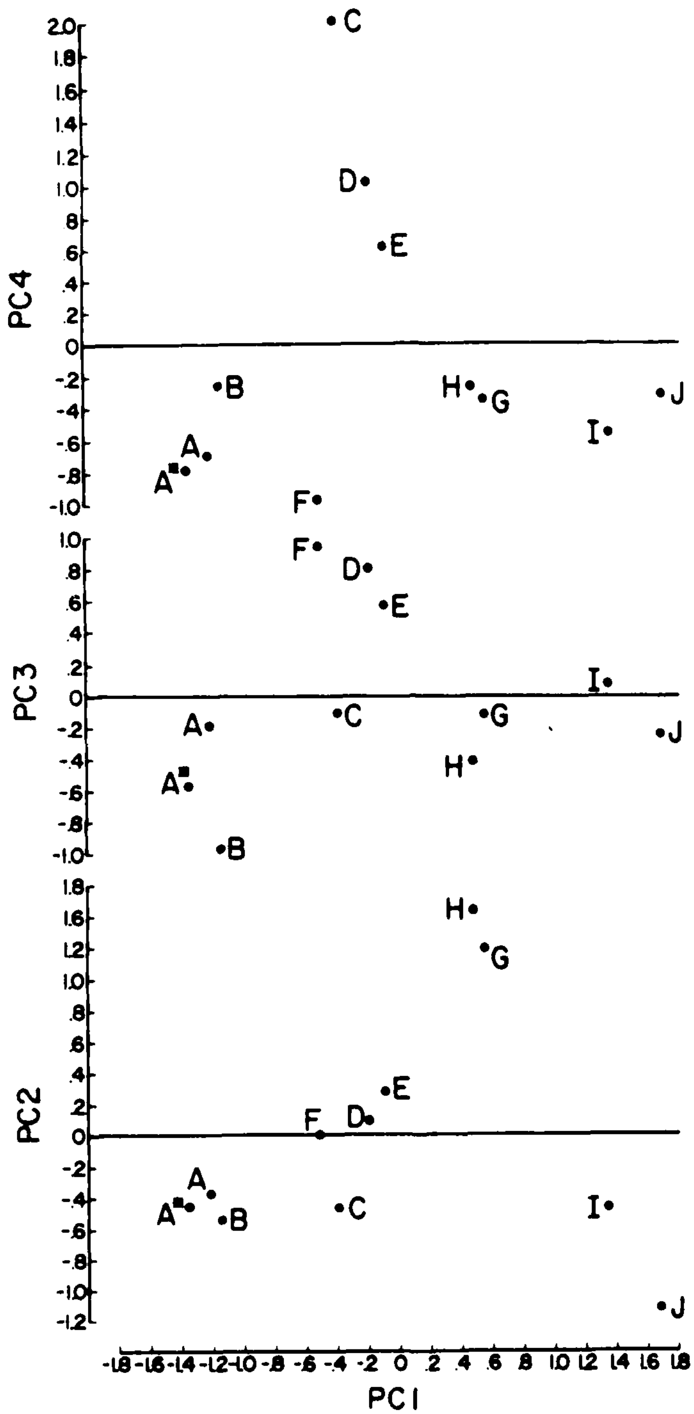


Figure 20. Ordination of Illinois study sites in principal components space based on the typical + tree guilds subset of habitat variables. (See Table 17. Study sites are identified in Appendix D. The loci represent mean principal components scores. A* represents study site A after the sapling edge was deleted from the analysis.)

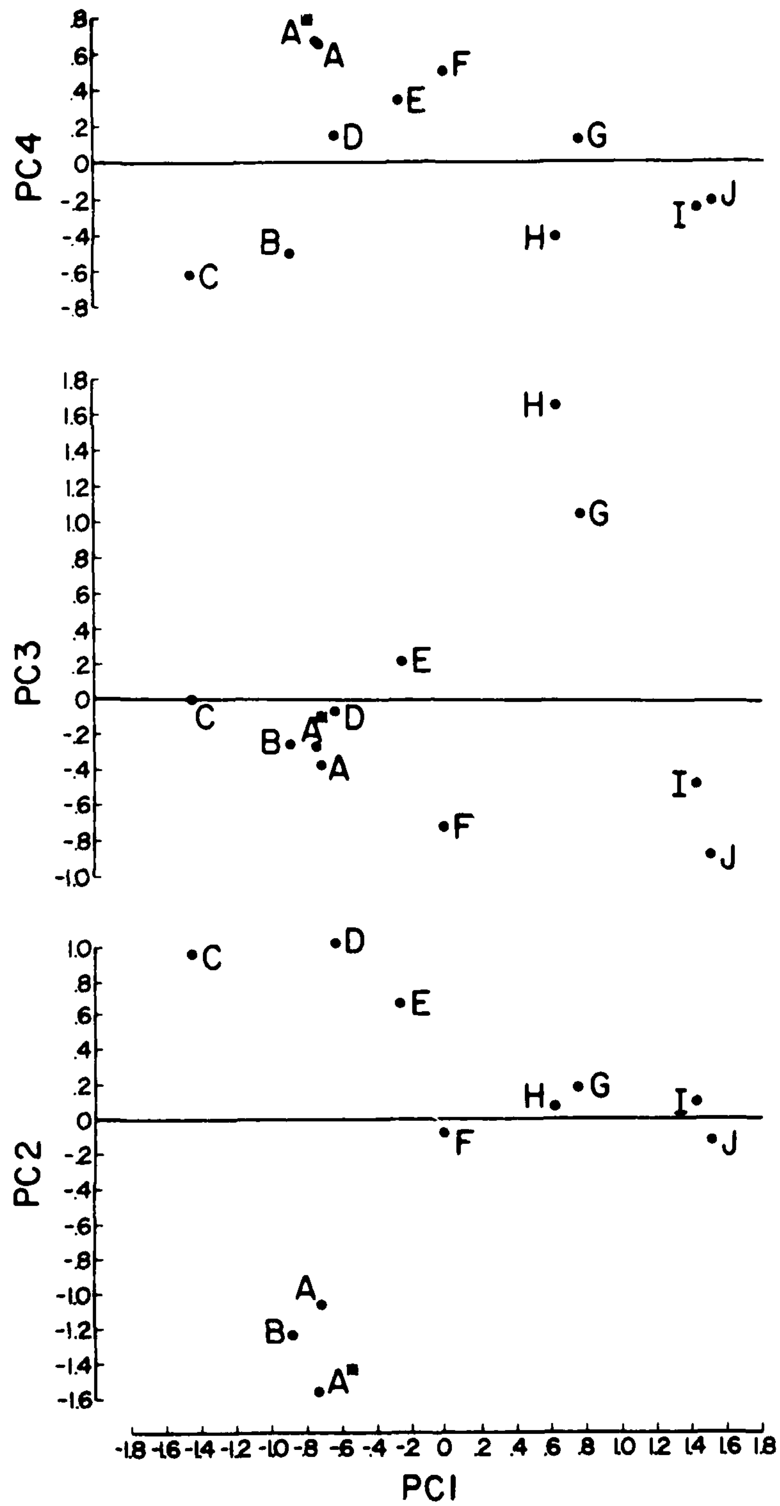


Figure 21. Ordination of Illinois study sites in principal components space based on the complete set of habitat variables. (See Table 18. Study sites are identified in Appendix D. The loci represent mean principal components scores. A* represents study site A after the sapling edge was deleted from the analysis.)

Table 15

Pearson Correlation of the First Four Principal Components With the Structural Subset of Illinois Habitat Variables, and Summary of the Important Associations of Habitat Variables With the First Four Principal Components

	PC 1	PC 2	PC 3	PC 4
L1			0.93	0.15[2]
L2	-0.20		0.32	0.67
L3	0.14[2]	0.64		0.38
L4	0.21	0.85		0.20
L5	0.20	0.86		
L6	0.42	0.76		
L7	0.71	0.43		
L8	0.86	0.23		0.13[2]
L9	0.92	0.19[1]		
L10	0.92	0.17[1]		
L11	0.84			
VEGVOL	0.65	0.49	0.48	0.15[2]
VHW	0.73	0.64		0.20
HZHW		0.26	-0.21	0.69
GC	-0.14[2]		0.91	
CC	0.62	0.67		
AQUA		-0.18[1]	-0.71	0.25
TOP				-0.36

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

PC 1 + Vegetation layers > 6.5 m especially 8.5 m
High vegetation volume
High canopy cover
High vertical heterogeneity

\- Vegetation layers 1-2 m
High ground cover

PC 2 + Vegetation layers 2 to 7.5 m
High vegetation volume
High canopy cover
High vertical heterogeneity
High horizontal heterogeneity

PC 3 + Vegetation layers < 2 m especially < 1 m
High ground cover
High vegetation volume

\- Presence of water

PC 4 + Vegetation layers 1 to 3 m
High horizontal heterogeneity

Table 16

Pearson Correlation of the First Four Principal Components With the Typical Subset of Illinois Habitat Variables, and Summary of the Important Associations of Habitat Variables With the First Four Principal Components

	PC 1	PC 2	PC 3	PC 4
GC	-0.21			0.76
CC	0.83	0.30		
SHRD		0.92		
SHRH	-0.16[2]	0.85		
SHRC	0.19[1]	0.90		
NSPSHR	0.46			0.71
CANAV	0.89		0.22	
CANMX	0.92		0.21	
NSPTRE	0.84	0.28		0.19[1]
SAP	0.49	0.55		
A	0.71	0.38		
B	0.86			
C	0.86			
D	0.77			
E	0.70	-0.19[1]	0.29	
F	0.41	-0.21	0.53	
G	0.22		0.77	
H			0.51	
I			0.59	0.15[2]
TREMAS	0.94		0.25	

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

PC 1 +
- High basal area of trees
- High canopy cover, average and maximum canopy heights
- High density of trees 8 to 53 cm dbh
- Many tree species

PC 1 −
- High ground cover
- High shrub heterogeneity

PC 2 +
- High shrub density, cover, and heterogeneity
- High density of trees < 15 cm dbh

PC 2 −
- High density of trees 38 to 68 cm dbh

PC 3 +
- High density of trees > 53 cm dbh

PC 4 +
- High ground cover
- Many shrub species

Table 17

Pearson Correlation of the First Four Principal Components with the Typical + Tree Guilds Illinois Habitat Variables, and Summary of the Important Associations of Habitat Variables With the First Four Principal Components

	PC 1	PC 2	PC 3	PC 4
GC	-0.35		0.26	
CC	0.76	0.30	0.35	0.26
SHRD	0.15[2]	-0.28	0.85	
SHRH	-0.18[1]	-0.16[2]	0.76	
SHRC	0.21	-0.24	0.84	
NSPSHR	0.28	0.21	0.39	-0.47
CANAV	0.87	0.31		
CANMX	0.84	0.38	0.17[1]	
NSPTRE	0.75	0.21	0.42	-0.25
SAP	0.47		0.56	0.35
A	0.59	0.30	0.51	0.26
B	0.74	0.32	0.24	-0.13[2]
C	0.80	0.21		-0.17[1]
D	0.79			-0.20
E	0.74	0.22	-0.21	-0.19[1]
F	0.33	0.59	-0.20	
G	0.18[1]	0.62		
H	0.15[2]	0.33	-0.15[2]	
I		0.25		
TREMAS	0.89	0.40	0.16[2]	
DUP	0.78	-0.44		-0.22
DBOT	0.26	0.80		
DRICH	0.65	-0.19[1]		-0.41
DEAR	0.14[2]		0.42	-0.50
DDEA	0.49	0.26		
DBL			0.24	0.88
DC		0.76		0.18[1]
DELM	0.37	0.52	0.14[2]	-0.37

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

PC 1 +
- High basal area of trees
- High canopy cover, average and maximum canopy height
- High density of trees 8 to 53 cm dbh
- Many tree species
- High dominance of upland tree species
- Snags

−
- High ground cover
- High shrub heterogeneity

PC 2 +
- High density of trees > 53 cm dbh
- High dominance of bottomland tree species

Table 17 (Cont'd)

	-	High dominance of upland tree species
PC 3	+	High shrub density, cover, and heterogeneity High density of trees < 15 cm dbh High dominance of early succession tree species
PC 4	+	High dominance of black locust
	-	High dominance of early succession and upland tree species Many shrub species

Table 18

Pearson Correlation of the First Four Principal Components With the Complete Set of Illinois Habitat Variables and Summary of the Important Associations of Habitat Variables With the First Four Principal Components

	PC 1	PC 2	PC 3	PC 4
L1				0.93
L2			-0.47	0.39
L3	0.29	0.56	-0.23	0.15[2]
L4	0.26	0.78		
L5	0.18[1]	0.77		
L6	0.31	0.76	0.15[1]	
L7	0.49	0.57	0.25	
L8	0.65	0.43	0.27	
L9	0.74	0.37	0.29	
L10	0.80	0 25	0.36	
L11	0.88		0.27	-0.13[2]
VEGVOL	0.55	0.55	0.17[1]	0.48
VHW	0.62	0.72	0.15[2]	
HZHW		0.45		
GC	-0.22			0.92
CC	0.54	0.73	0.21	
AQUA			0.13[2]	-0.59
TOP		-0.20		
SHRD	0.17[1]	0.48	-0.62	0.21
SHRH		0.36	-0.48	0.17[1]
SHRC	0.24	0.50	-0.58	0.21
NSPSHR	0.51			0.38
CANAV	0.76	0.42	0.31	
CANMX	0.78	0.42	0.31	
NSPTRE	0.78	0.43		
SAP	0.21	0.77		
A	0.39	0.75	0.14[2]	
B	0.73	0.38	0.20	
C	0.80	0.19[1]	0.18[1]	
D	0.80			
E	0.74		0.26	
F	0.39		0.56	
G	0.25		0.53	
H	0.17[1]		0.33	
I			0.19[1]	

Table 18 (Cont'd)

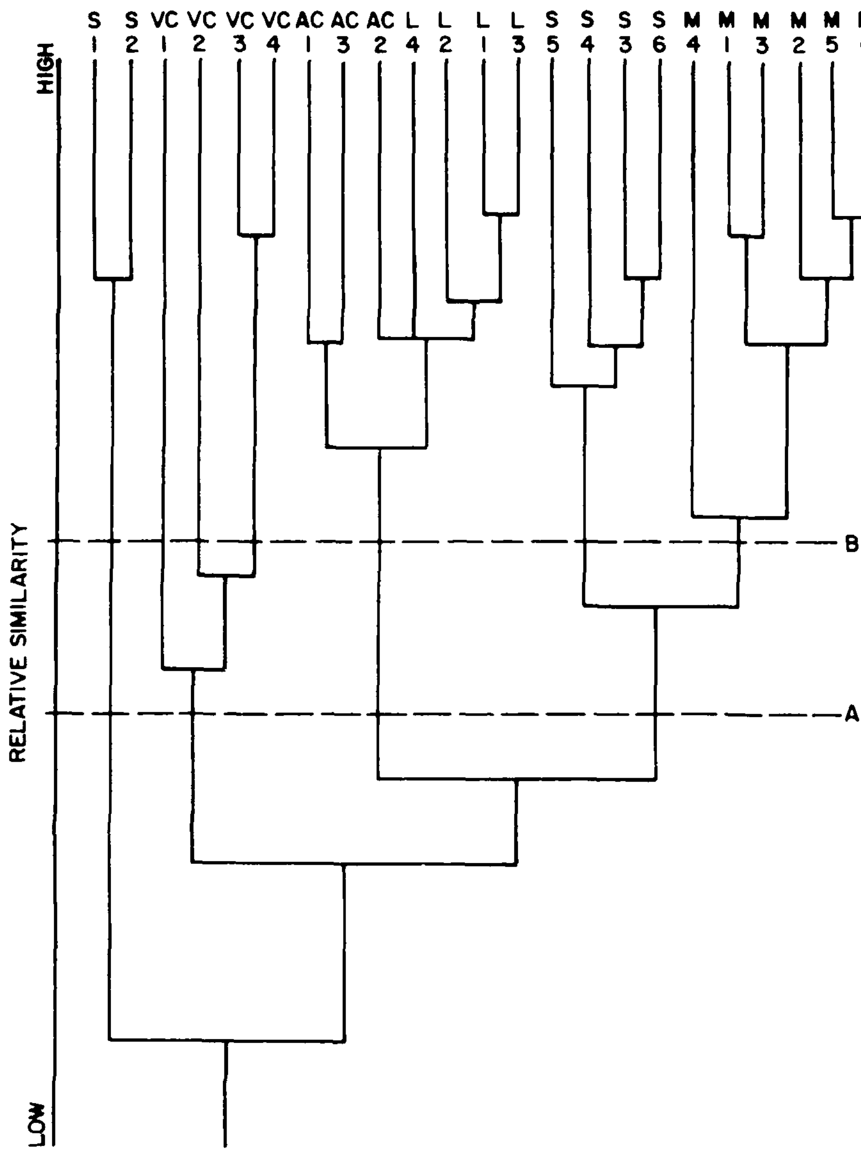
	PC 1	PC 2	PC 3	PC 4
TREMAS	0.77	0.48	0.32	
DUP	0.75		-0.38	
DBOT	0.32		0.72	
DRICH	0.72		-0.24	
DEAR	0.34	0.17[1]	-0.16[2]	0.18[1]
DDEA	0.47	0.19[1]	0.18[1]	
DBL	-0.51	0.60		
DC		0.20	0.69	
DELM	0.52	0.17[1]	0.34	

$P \leq 0.0001$

1 $0.0001 < P \leq 0.001$

2 $0.001 < P \leq 0.01$

PC 1	+	Vegetation layers > 6.5 m, especially > 8.5 m High vegetation volume and basal area of trees High canopy cover, average and maximum canopy heights High vertical heterogeneity High density of trees 15 to 53 cm dbh Many tree species High dominance of upland tree species Snags
	-	High ground cover High dominance of black locust
PC 2	+	Vegetation layers 2 to 7.5 m High vegetation volume High canopy cover High vertical heterogeneity High density of trees < 15 cm dbh High dominance of black locust High shrub density, cover, and heterogeneity High horizontal heterogeneity of vegetation
PC 3	+	Vegetation layers > 6.5 m Scarcity of vegetation layers < 5 m High density of trees > 53 cm dbh High dominance of bottomland tree species
	-	Vegetation layers 1 to 3 m High dominance of upland and early succession tree species High shrub density, cover, and heterogeneity
PC 4	+	Vegetation layers < 2 m, especially < 1 m High ground cover High dominance of early succession tree species High vegetation volume
	-	Presence of aquatic habitats



POSSIBLE GROUPS

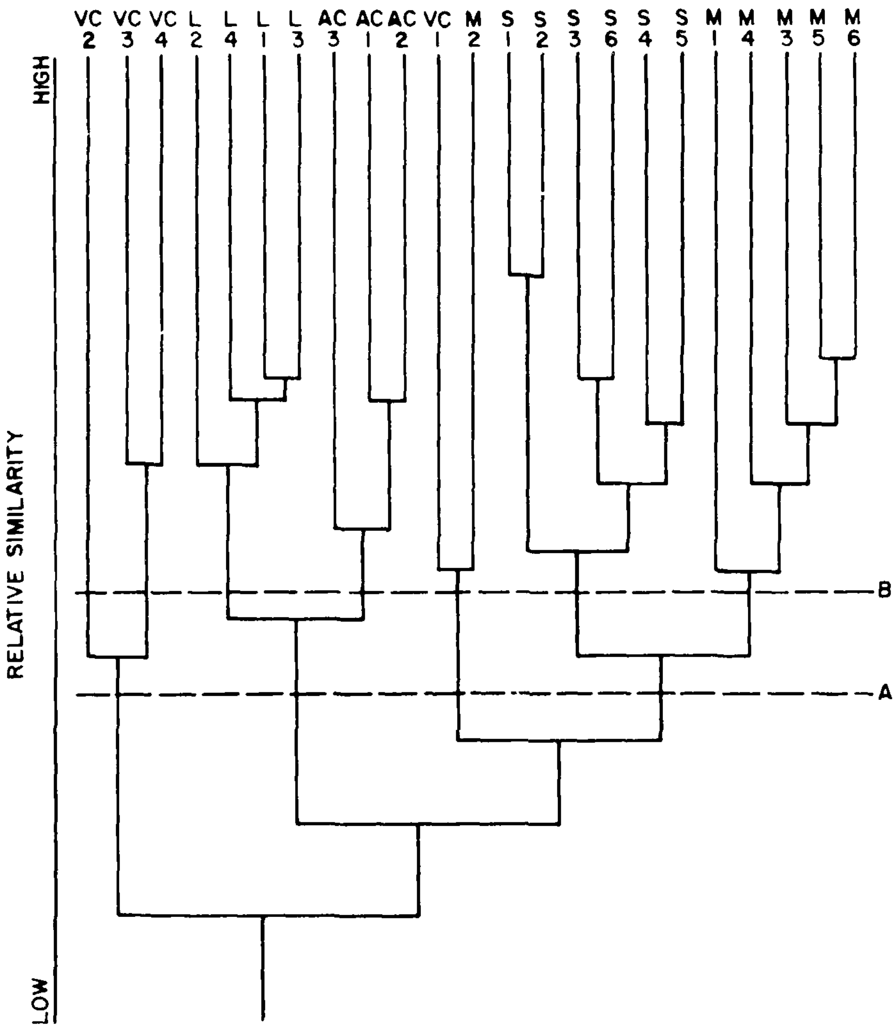
A

1) S1, S2
2) VC1, VC2, VC3, VC4
3) AC1, AC3, AC2, L4, L2, L1, L3
4) S5, S4, S3, S6, M4, M1, M3, M2, M5, M6

B

1) S1, S2
2) VC1
3) VC2
4) VC3, VC4
5) AC1, AC3, AC2, L4, L2, L1, L3
6) S5, S4, S3, S6
7) M4, M1, M3, M2, M5, M6

Figure 22. Cluster analysis of the Mojave Desert transects using the habitat variable subset that was optimal for CAD. (The horizontal lines indicate two possible groups of clusters.)



POSSIBLE GROUPS

A

1) VC2, VC3, VC4
2) L2, L4, L1, L3, AC3, AC1, AC2
3) VC1, M2
4) S1, S2, S3, S6, S4, S5, M1, M4, M3, M5, M6

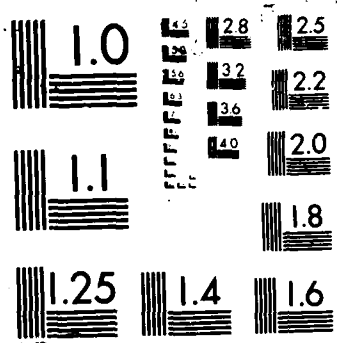
B

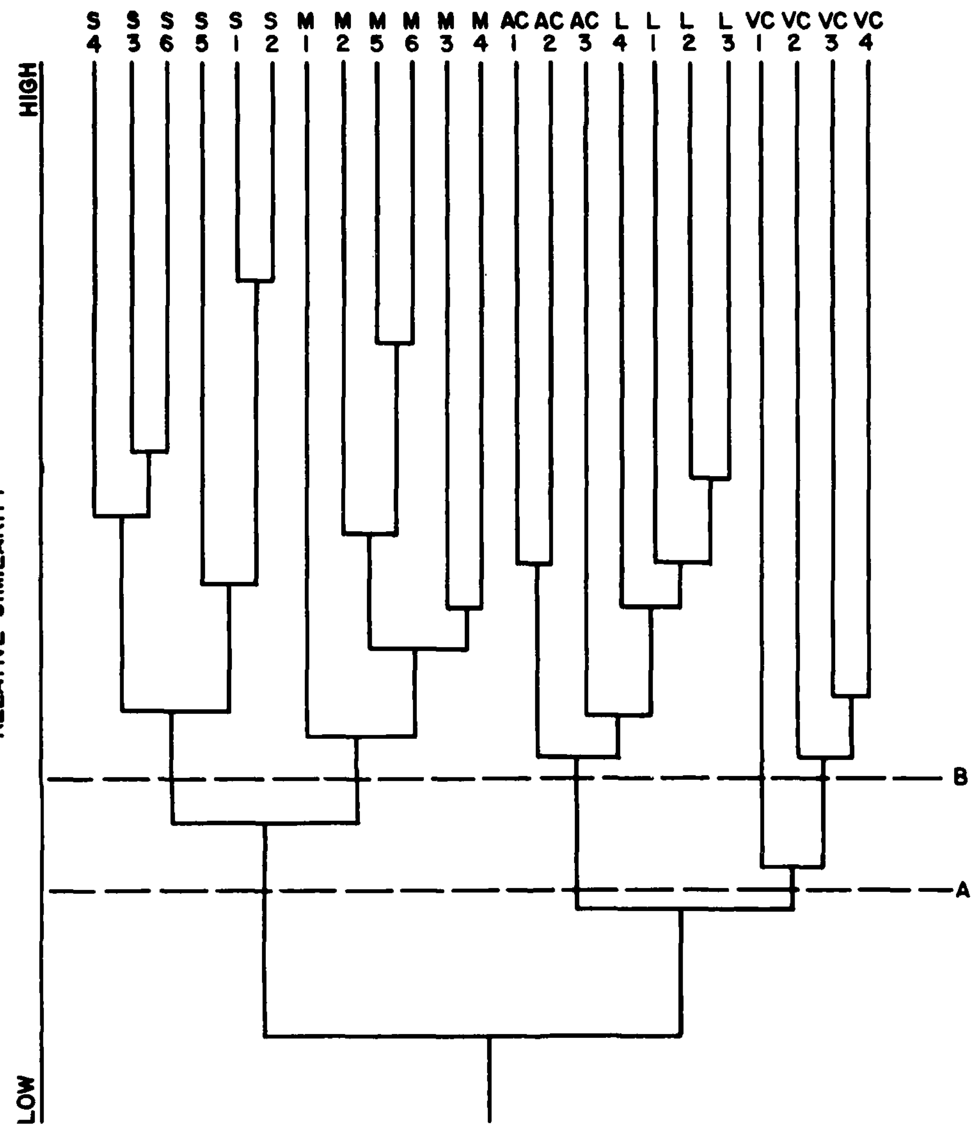
1) VC2
2) VC3, VC4
3) L2, L4, L1, L3
4) AC3, AC1, AC2
5) VC1, M2
6) S1, S2, S3, S6, S4, S5
7) M1, M4, M3, M5, M6

Figure 23. Cluster analysis of the Mojave Dese[illegible] variable subset that wa[illegible] indicate two possible gr[illegible]

AD-A187 294 ENVIRONMENTAL GRADIENT ANALYSIS ORDINATION AND CLASSIFICATION IN ENVIRONM..(U) CONSTRUCTION ENGINEERING RESEARCH LAB (ARMY) CHAMPAIGN IL 2/2

UNCLASSIFIED A J KRZYSIK SEP 87 CERL-TR-N-87/19 F/G 6/6 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6



POSSIBLE GROUPS

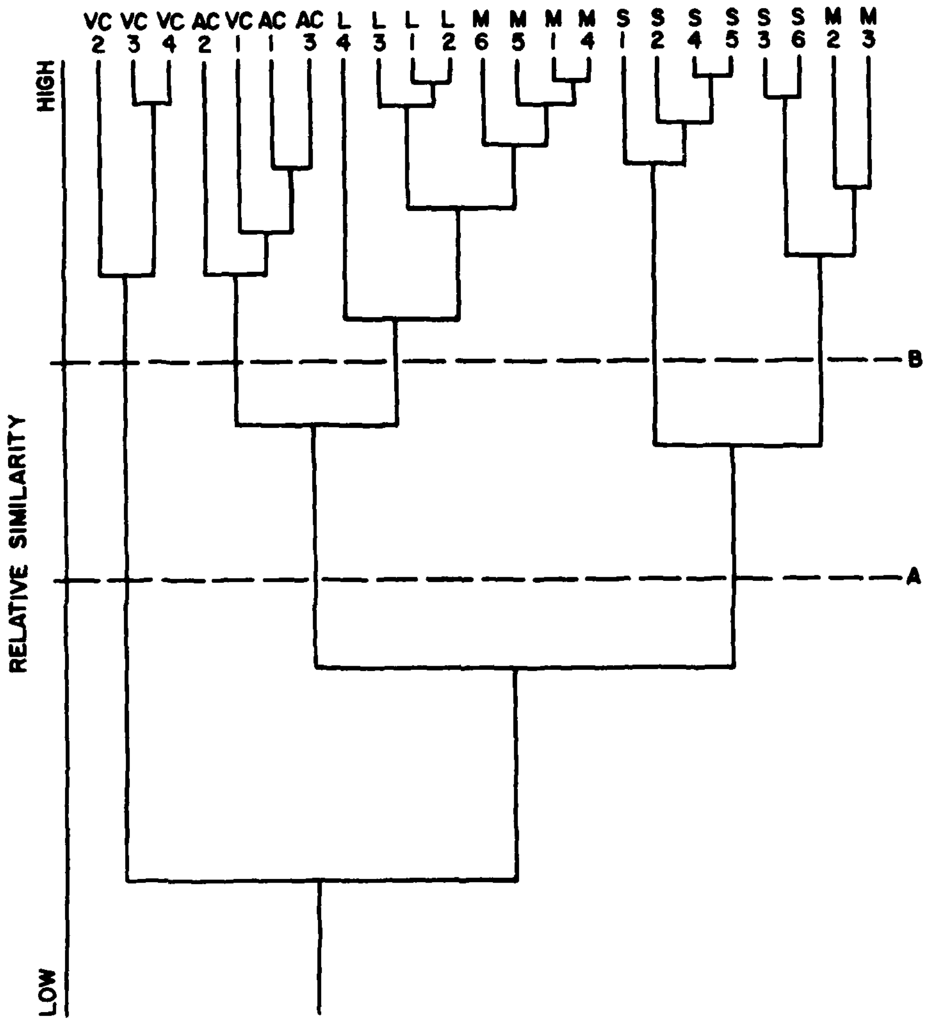
A

1) S4, S3, S6, S5, S1, S2, M1, M2, M5, M6, M3, M4
2) AC1, AC2, AC3, L4, L1, L2, L3
3) VC1, VC2, VC3, VC4

B

1) S4, S3, S6, S5, S1, S2
2) M1, M2, M5, M6, M3, M4
3) AC1, AC2, AC3, L4, L1, L2, L3
4) VC1
5) VC2, VC3, VC4

Figure 24. Cluster analysis of the Mojave Desert transects using the complete set of habitat variables. (The horizontal lines indicate two possible groups of clusters.)



POSSIBLE GROUPS

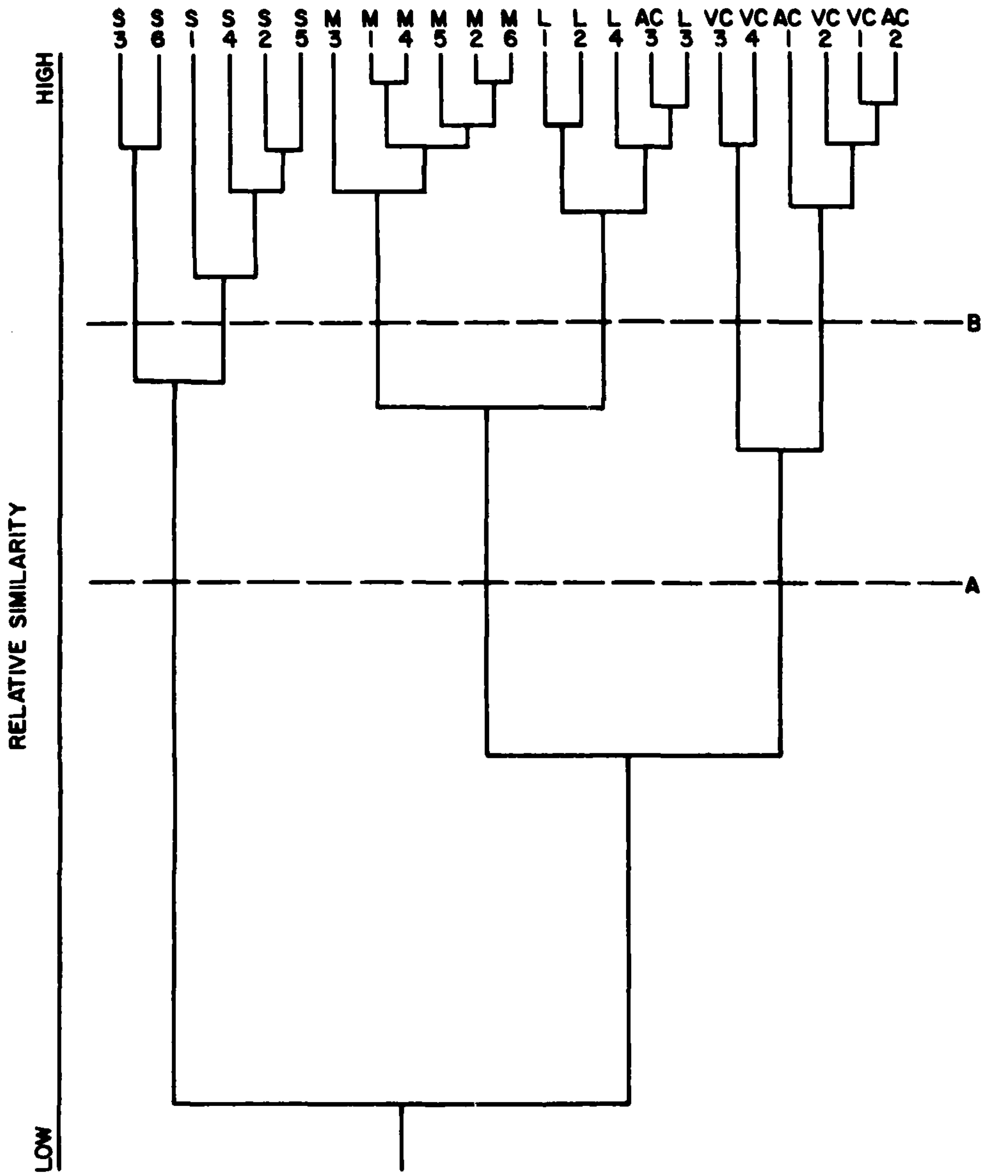
A

1) VC2,VC3,VC4
2) AC2,VC1, AC1,AC3,L4,L3,L1,L2,M6,M5,M1,M4
3) S1,S2,S4,S5,S3,S6,M2,M3

B

1) VC2,VC3,VC4
2) AC2,VC1,AC1,AC3
3) L4,L3,L1,L2,M6,M5,M1,M4
4) S1,S2,S4,S5
5) S3,S6,M2,M3

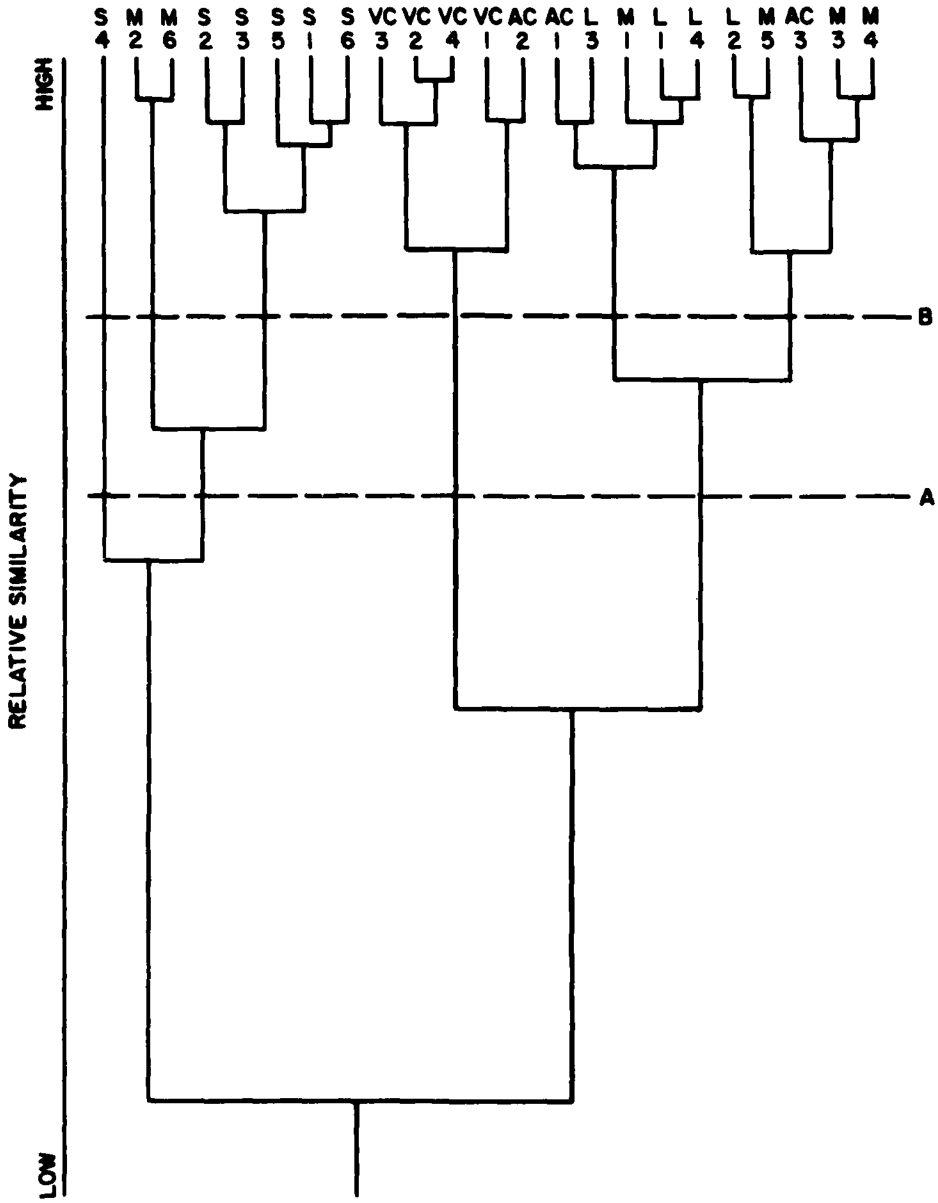
Figure 27. Cluster analysis of the Mojave Desert transects using the first two principal components. (The horizontal lines indicate two possible groups of clusters.)



POSSIBLE GROUPS

A	B
1) S3,S6,S1,S4,S2,S5	1) S3,S6
2) M3,M1,M4,M5,M2,M6,L1,L2,L4,AC3,L3	2) S1,S4,S2,S5
3) VC3,VC4,AC1,VC2,VC1,AC2	3) M3,M1,M4,M5,M2,M6
	4) L1,L2,L4,AC3,L3
	5) VC3,VC4
	6) AC1,VC2,VC1,AC2

Figure 28. Cluster analysis of the Mojave Desert transects using the first principal component. (The horizontal lines indicate two possible groups of clusters.)



POSSIBLE GROUPS

A	B
1) S4	1) S4
2) M2,M6,S2,S3,S5,S1,S6	2) M2,M6
3) VC3,VC2,VC4,VC1,AC2	3) S2,S3,S5,S1,S6
4) AC1,L3,M1,L1,L4,L2,M5,AC3,M3,M4	4) VC3,VC2,VC4,VC1,AC2
	5) AC1,L3,M1,L1,L4
	6) L2,M5,AC3,M3,M4

Figure 29. Cluster analysis of the Mojave Desert transects using the first principal component from a PCA where the maximum number of eigenvectors was extracted.

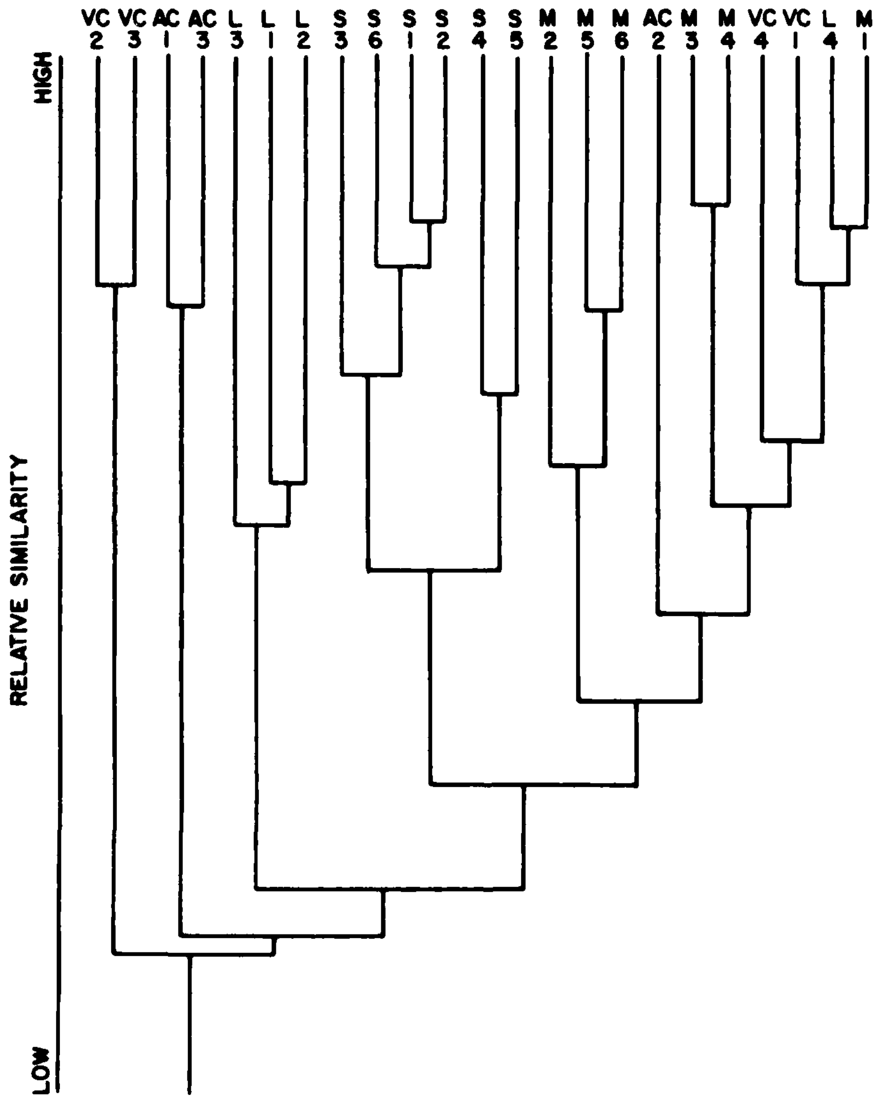


Figure 30. Cluster analysis of the Mojave Desert transects using the first four principal components from a PCA where the maximum number of eigenvectors was extracted.

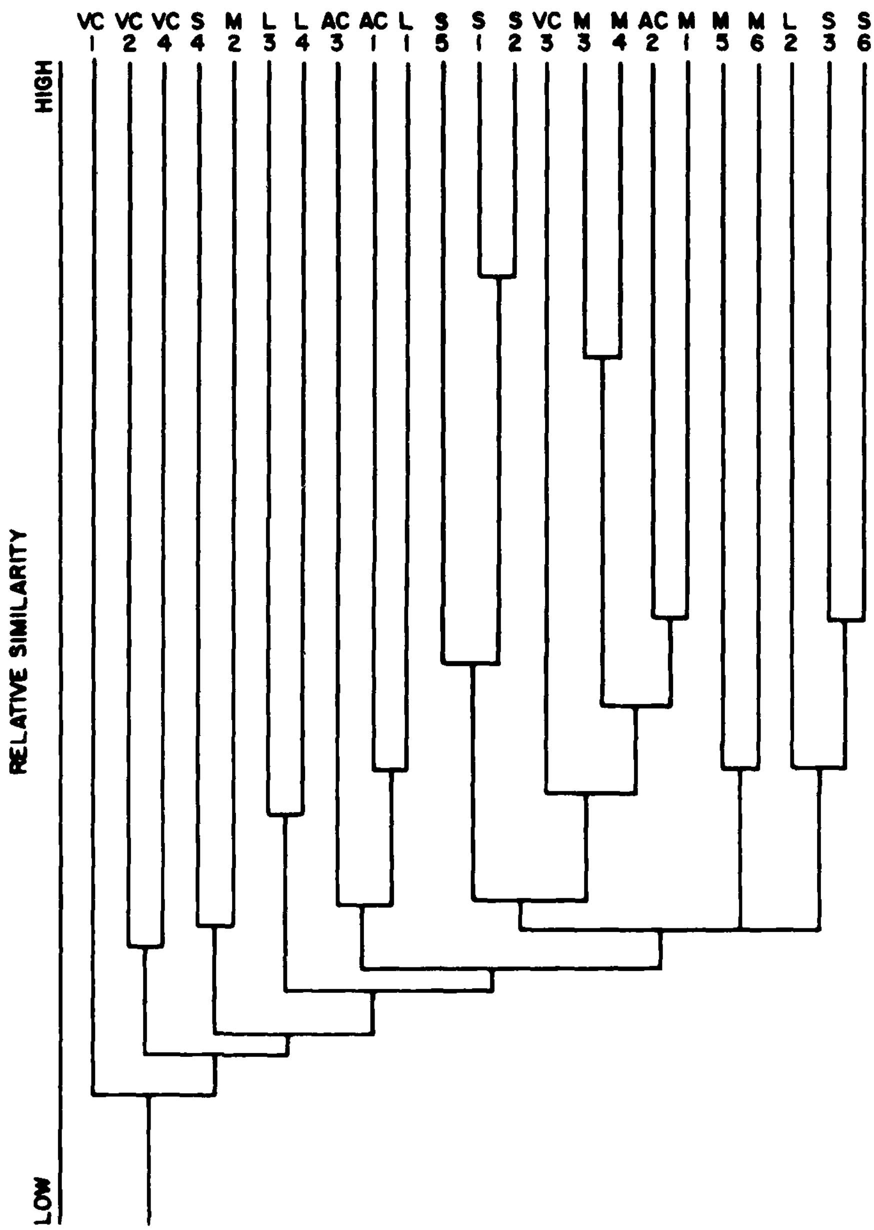


Figure 31. Cluster analysis of the Mojave Desert transects using the first 10 principal components from a PCA where the maximum number of eigenvectors was extracted.

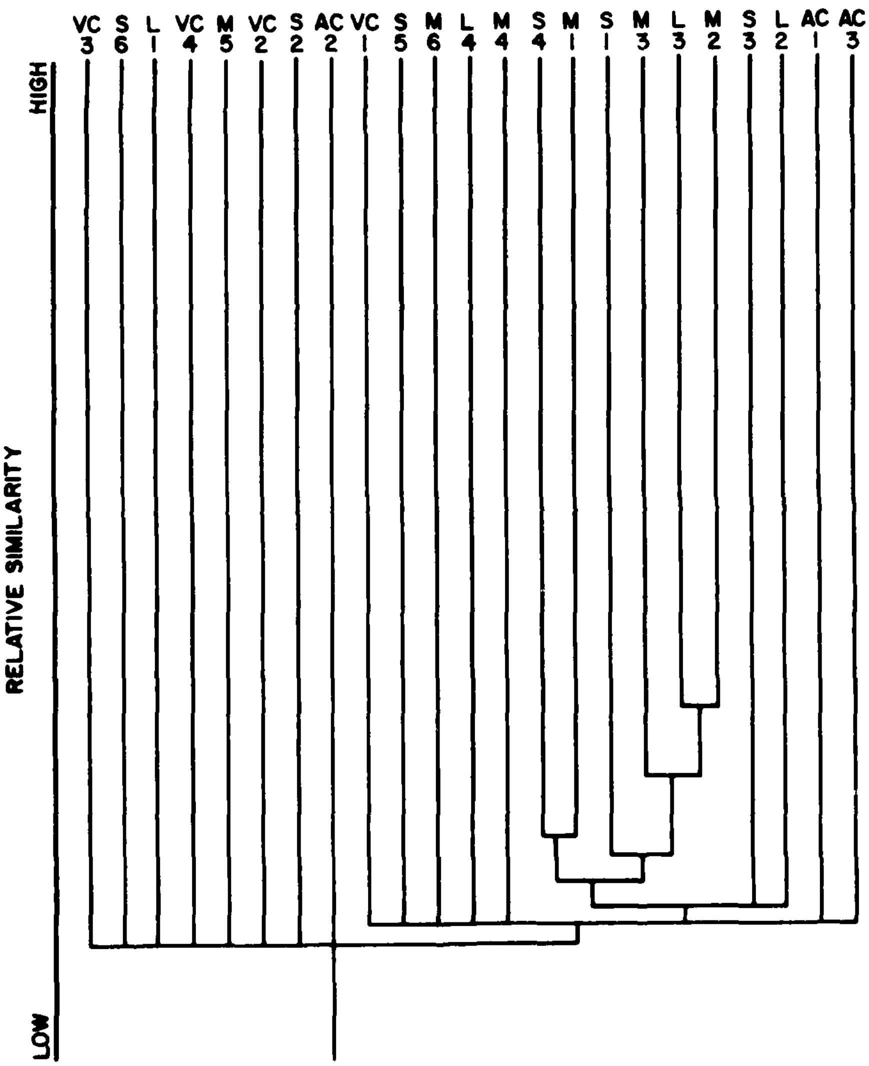


Figure 32. Cluster analysis of the Mojave Desert transects using the first 20 principal components from a PCA where the maximum number of eigenvectors was extracted.

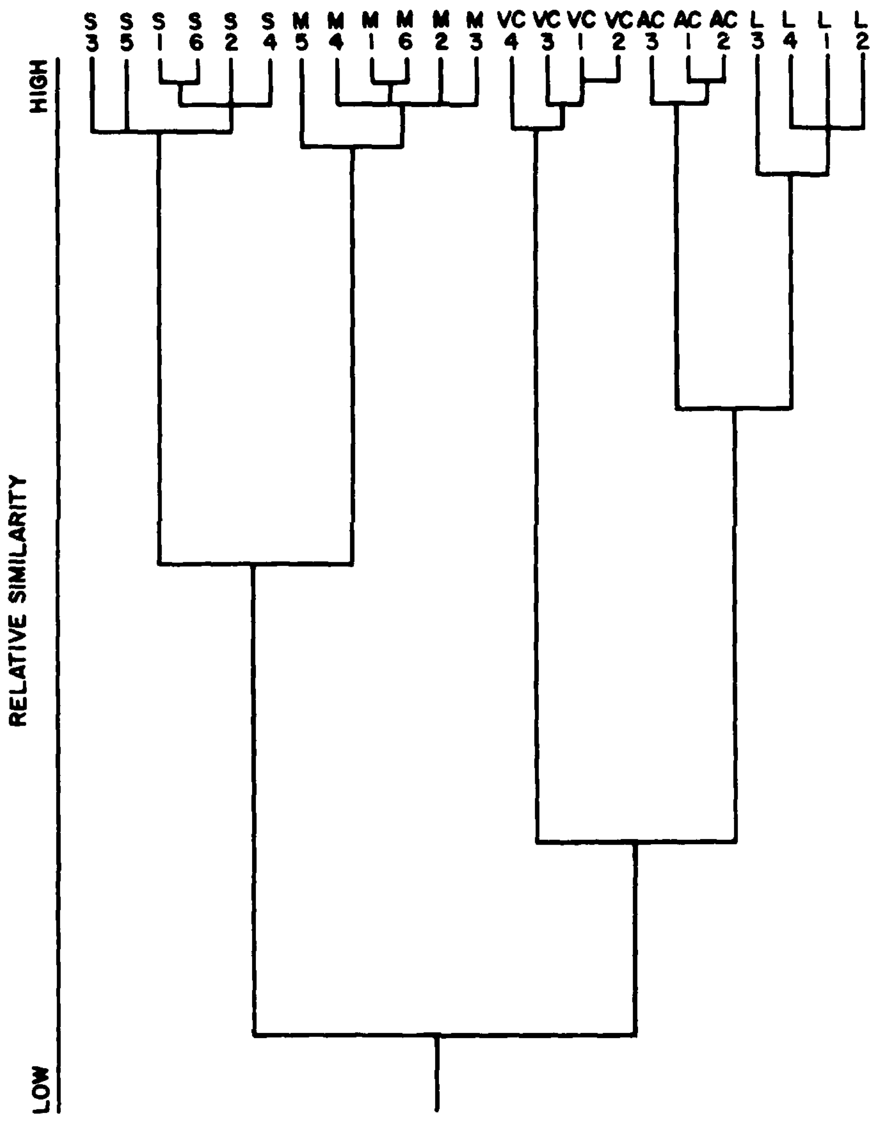


Figure 33. Cluster analysis of the Mojave Desert transects using the first four canonical variates.

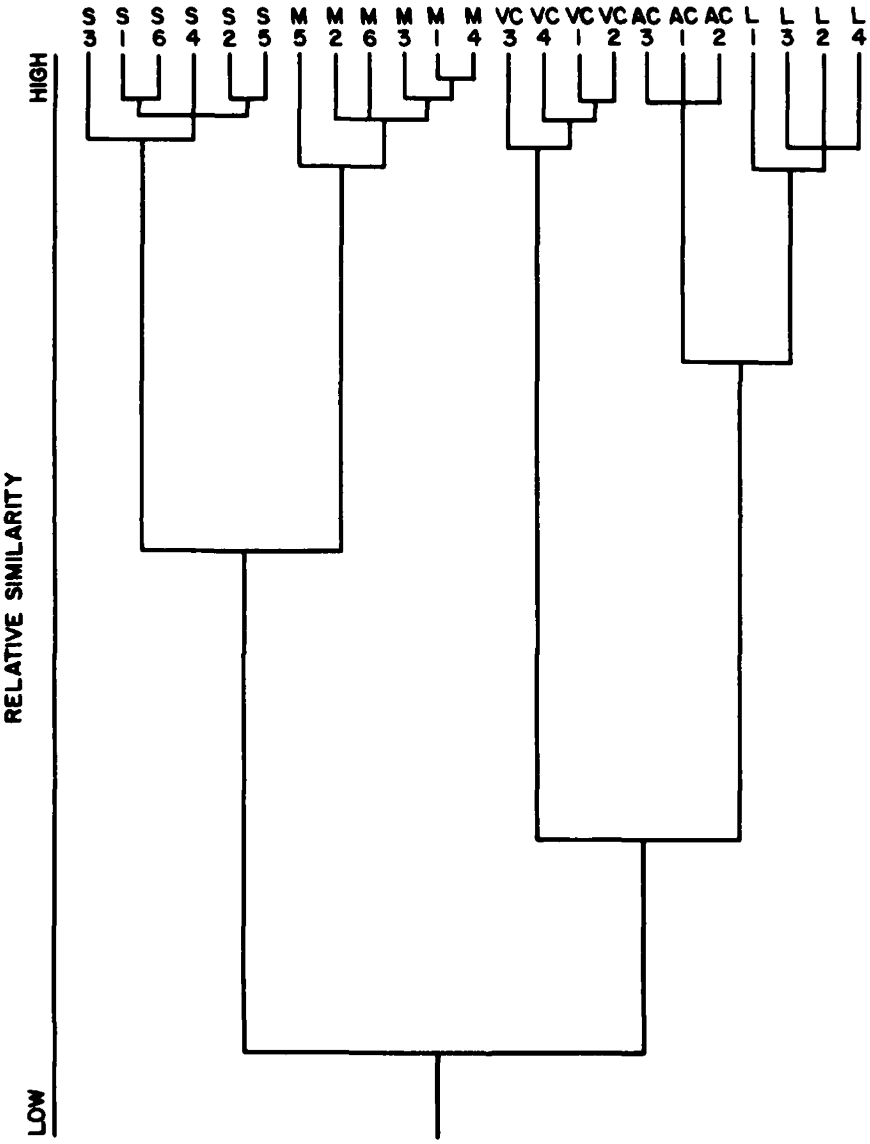


Figure 34. Cluster analysis of the Mojave Desert transects using the first three canonical variates.

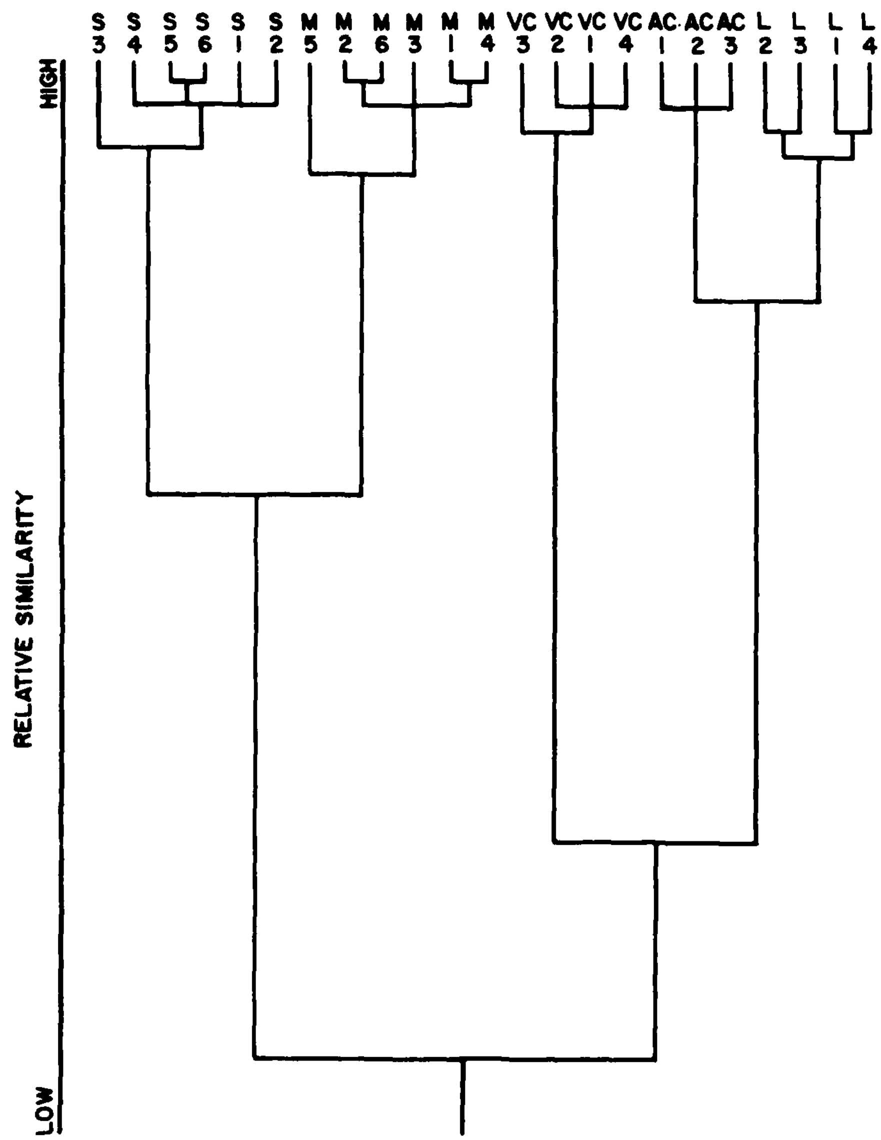


Figure 35. Cluster analysis of the Mojave Desert transects using the first two canonical variates.

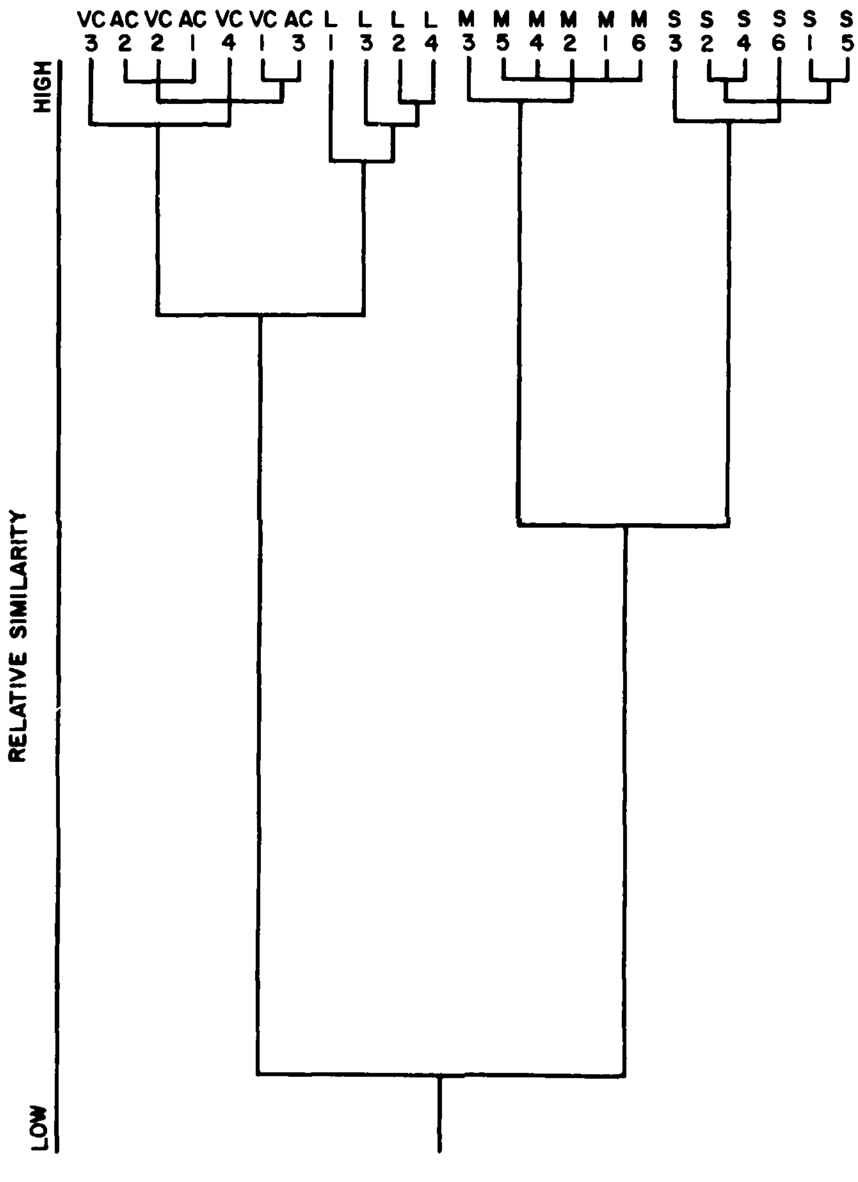


Figure 36. Cluster analysis of the Mojave Desert transects using the first canonical variate.

REFERENCES

1. Anderberg, M. R., Cluster Analysis for Applications (Academic Press, New York, NY, 1973).

2. Anderson, S. H., and H. H. Shugart, Jr., "Habitat Selection of Breeding Birds in an East Tennessee Deciduous Forest," Ecology, Vol 55 (1974), pp 828-837.

3. Anderson, T. W., "Asymptotic Theory for Principal Component Analysis," Ann. Math. Stat., Vol 34 (1963), pp 122-148.

4. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, 2nd ed. (John Wiley & Sons, New York, NY, 1984).

5. Austin, M. P., and I. Noy-Meir, "The Problem of Nonlinearity in Ordination: Experiments With Two-Gradient Models," J. Ecol., Vol 59 (1971), pp 763-774.

6. Balda, R. P., "Foliage Use by Birds of the Oak-Juniper Woodland and Ponderosa Pine Forest in South-Eastern Arizona," Condor, Vol 71 (1969), pp 399-412.

7. Baroni-Urbani, C., and M. W. Buser, "Similarity of Binary Data," Syst. Zool., Vol 25 (1976), pp 251-259.

8. Berenbaum, M., and P. Feeny, "Toxicity of Angular Furanocoumarins to Swallowtail Butterflies: Escalation in a Coevolutionary Arms Race," Science, Vol 212 (1981), pp 927-929.

9. Boesch, D. F., Application of Numerical Classification in Ecological Investigations of Water Pollution, Report No. EPA-600/3-77-033 (U.S. Environmental Protection Agency, Corvallis, OR, 1977).

10. Brower, L. P., and J. V. Z. Brower, "Birds, Butterflies, and Plant Poisons: A Study in Ecological Chemistry," Zoologica, Vol 49 (1964), pp 137-159.

11. Carleton, T. J., "Non-Centered Component Analysis of Vegetation Data: A Comparison of Orthogonal and Oblique Rotation," Vegetatio, Vol 42 (1980), pp 59-66.

12. Carroll, J. B., "Approximating Simple Structure in Factor Analysis," Psych., Vol 18 (1953), pp 23-38.

13. Cassie, R. M., and A. D. Michael, "Fauna and Sediments of an Intertidal Mud Flat: A Multivariate Analysis," J. Exper. Mar. Biol. Ecol., Vol 2 (1968), pp 1-23.

14. Cheetham, A. H., and J. E. Hazel, "Binary (Presence-Absence) Similarity Coefficients," J. Paleontology, Vol 43 (1969), pp 1130-1136.

15. Clifford, H. T., and W. Stephenson, An Introduction to Numerical Classification (Academic Press, New York, NY, 1975).

16. Crawford, H. S., R. C. Hooper, and R. W. Titterington, "Songbird Population Responses to Silvicultural Practices in Central Appalachian Hardwood," J. Wildl. Manage., Vol 45 (1981), pp 680-692.

17. Dillon, W. R., and M. Goldstein, Multivariate Analysis: Methods and Applications (John Wiley & Sons, New York, NY, 1984).

18. Dixon, W. J. (ed.), BMDP Statistical Software (Univ. California Press, Berkeley, CA, 1983 printing with additions).

19. Doran, J. E., and F. R. Hodson, Mathematics and Computers in Archeology (Harvard Univ. Press, Cambridge, MA, 1975).

20. Dowdy, S., and S. Wearden, Statistics for Research (John Wiley & Sons, New York, NY, 1983).

20a. Draper, N. R., and H. Smith, Applied Regression Analysis, 2nd ed. (John Wiley & Sons, New York, NY, 1981).

21. Dunn, G., and B. S. Everitt, An Introduction to Mathematical Taxonomy (Cambridge Univ. Press, Cambridge, England, 1982).

22. Durbin, J., "Weak Convergences of the Sample Distribution Function When Parameters Are Estimated," J. Amer. Stat. Assoc., Vol 1 (1973), pp 279-290.

23. Ebert, T. A., and G. S. McMaster, "Regular Pattern of Desert Shrubs: A Sampling Artifact?," J. Ecol., Vol 69 (1981), pp 559-564.

24. Ehrlich, P. R., and P. H. Raven, "Butterflies and Plants: A Study in Coevolution," Evolution, Vol 18 (1964), pp 586-608.

25. Eisenbeis, R. A., G. G. Gilbert, and R. B. Avery, "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis," Commun. Stat., Vol 2 (1973), pp 205-219.

26. Everitt, B., Cluster Analysis, 2nd ed. (Halsted Press, New York, NY, 1980).

27. Feeny, P., "Biochemical Coevolution Between Plants and Their Insect Herbivores," in L. E. Gilbert and P. H. Raven, Eds., Coevolution of Animals and Plants (Univ. Texas Press, Austin, TX, 1975), pp 3-19.

28. Feeny, P., "Plant Apparency and Chemical Defense," Rec. Adv. Phytochem., Vol 10 (1976), pp 1-40.

29. Feeny, P., "Defensive Ecology of the Cruciferae," Ann. Missouri Bot. Gard., Vol 64 (1977), pp 221-234.

30. Ferguson, G. A., "The Concept of Parsimony in Factor Analysis," Psychol., Vol 19 (1954), pp 281-290.

31. Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, Vol 7 (1936), pp 179-188.

32. Franzreb, K. E., "Tree Species Used by Birds in Logged and Unlogged Mixed Coniferous Forests," Wilson Bull., Vol 90 (1978), pp 221-238.

33. Gauch, H. G., Jr., Multivariate Analysis in Community Ecology (Cambridge Univ. Press, Cambridge, England, 1982).

34. Gauch, H. G., Jr., and R. H. Whittaker, "Coenocline Simulation," Ecology, Vol 53 (1972), pp 446-451.

35. Gauch, H. G., Jr., and R. H. Whittaker, "Comparison of Ordination Techniques," Ecology, Vol 53 (1972), pp 868-875.

36. Gauch, H. G., Jr., R. H. Whittaker, and S. B. Singer, "A Comparative Study of Nonmetric Ordinations," J. Ecol., Vol 69 (1981), pp 135-152.

37. Gauch, H. G., Jr., R. H. Whittaker, and T. R. Wentworth, "A Comparative Study of Reciprocal Averaging and Other Ordination Techniques," J. Ecol., Vol 65 (1977), pp 157-174.

38. Gilbert, L. E., "Development of Theory in the Analysis of Insect-Plant Interactions," in D. J. Horn, G. P. Stairs, and R. D. Mitchell, eds., Analysis of Ecological Systems (Ohio State Univ. Press, Columbus, OH, 1979), pp 117-154.

39. Gnanadesikan, R., Methods for Statistical Data Analysis of Multivariate Observations (John Wiley & Sons, New York, NY, 1977).

39a. Gower, J. C., "A Comparison of Some Methods of Cluster Analysis," Biometrics, Vol 23 (1967), pp 623-637.

40. Green, R. H., Sampling Design and Statistical Methods for Environmental Biologists (John Wiley & Sons, New York, NY, 1979).

41. Greig-Smith, P., "The Development of Numerical Classification and Ordination," Vegetatio, Vol 42 (1980), pp 1-9.

42. Habbema, J. D. F., and J. Hermans, "Selection of Variables in Discriminant Analysis by F-Statistic and Error Rate," Technometrics, Vol 19 (1977), pp 487-493.

43. Harman, H. H., Modern Factor Analysis, 3rd ed. (Univ. Chicago Press, Chicago, IL, 1976).

44. Harris, R. J., A Primer of Multivariate Statistics (Academic Press, New York, NY, 1975).

45. Hartigan, J. A., Clustering Algorithms (John Wiley & Sons, New York, NY, 1975).

46. Hartley, P. H. T., "An Ecological Study of the Feeding Habits of the English Titmice," J. Anim. Ecol., Vol 22 (1953), pp 261-288.

47. Hawkins, D. M., "A New Test for Multivariate Normality and Homoscedasticity," Technometrics, Vol 23 (1981), pp 105-110.

48. Hill, M. O., "Correspondence Analysis: A Neglected Multivariate Method," J. Royal Stat. Soc., Series C, Vol 23 (1974), pp 340-354.

49. Hill, M. O., and H. G. Gauch, Jr., "Detrended Correspondence Analysis: An Improved Ordination Technique," Vegetatio, Vol 42 (1980), pp 47-58.

50. Holmes, R. T., R. E. Bonney, Jr., and S. W. Pacala, "Guild Structure of the Hubbard Brook Bird Community: A Multivariate Approach," Ecology, Vol 60 (1979), pp 512-520.

51. Holmes, R. T., and S. K. Robinson, "Tree Species Preferences of Foraging Insectivorous Birds in a Northern Hardwood Forest," Oecologia, Vol 48 (1981), pp 31-35.

52. James, F. C., "Ordinations of Habitat Relationships Among Breeding Birds," Wilson Bull., Vol 83 (1971), pp 215-236.

53. James, F. C., and H. H. Shugart, "A Quantitative Method of Habitat Description," Audubon Field Notes, Vol 24 (1970), pp 727-736.

54. James, R. D., "The Comparative Foraging Behaviour of Yellow-Throated and Solitary Vireos: The Effect of Habitat and Sympatry," in J. G. Dickson, R. N. Connor, R. R. Flect, J. A. Jackson, and J. C. Kroll, Eds., The Role of Insectivorous Birds in Forest Ecosystems (Academic Press, New York, NY, 1979), pp 137-163.

55. Jackson, J. A., "A Quantitative Study of the Foraging Ecology of Downy Woodpeckers," Ecology, Vol 51 (1970), pp 318-323.

56. Jackson, J. A., "Tree Surfaces as Foraging Substrates for Insectivorous Birds," in J. G. Dickson, R. N. Connor, R. H. Fleet, J. A. Jackson, and J. C. Kroll, Eds., The Role of Insectivorous Birds in Forest Ecosystems (Academic Press, New York, NY, 1979), pp 69-93.

57. Jermy, T., "Evolution of Insect/Host Plant Relationships," Amer. Natur., Vol 124 (1984), pp 609-630.

58. Johnston, R. F., "Replication of Habitat Profiles for Birds," Occ. Pap. Mus. Nat. Hist., No. 82 (Univ. Kansas, Lawrence, KS, 1979), pp 1-11.

59. Kaiser, H. F., "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, Vol 23 (1958), pp 187-200.

60. Karr, J. R., and R. R. Roth., "Vegetation Structure and Avian Diversity in Several New World Areas," Amer. Natur., Vol 105 (1971), pp 423-435.

61. King, T. J., and S. R. J. Woodell, "Are Regular Patterns in Desert Shrubs Artefacts of Sampling?," J. Ecol., Vol 72 (1984), pp 295-298.

62. Krzanowski, W. J., "The Performance of Fisher's Linear Discriminant Function Under Non-Optimal Conditions," Technometrics, Vol 19 (1977), pp 191-200.

63. Krzysik, A. J., "Habitat Relationships and the Effects of Environmental Impacts on the Bird and Small Mammal Communities of the Central Mojave Desert," in W. C. McComb, Ed., Proceedings--Workshop on Management of Nongame Species and Ecological Communities (Univ. Kentucky, Lexington, KY, 1984), pp 358-394.

64. Krzysik, A. J., Ecological Assessment of the Effects of Army Training Activities on a Desert Ecosystem: National Training Center, Fort Irwin, California, Technical Report N-85/13/ADA159248 (U.S. Army Construction Engineering Research Laboratory, 1985).

65. Kshirsagar, A. M., Multivariate Analysis (Marcel Dekker, New York, NY, 1972).

66. Lachenbruch, P. A., C. Sneeringer, and L. T. Revo, "Robustness of the Linear and Quadratic Discriminant Functions to Certain Types of Non-Normality," Commun. Stat., Vol 1 (1973), pp 39-56.

66a. Lance, G. N., and W. T. Williams, "A Generalized Sorting Strategy for Computer Classifications," Nature, Vol 212 (1966), p 218.

66b. Lance, G. N., and W. T. Williams, "A General Theory of Classificatory Sorting Strategies. I. Hierarchical Systems." Computer J., Vol 9 (1967), pp 373-380.

67. Legendre, L., and P. Legendre, Numerical Ecology (Elsevier, New York, NY, 1983).

68. Levin, D. A., "The Chemical Defenses of Plants to Pathogens and Herbivores," Ann. Rev. Ecol. Syst., Vol 7 (1976), pp 121-159.

69. Lovejoy, T. E., "Bird Species Diversity and Composition in Amazonian Rain Forests," Amer. Zool., Vol 12 (1972), pp 711-712.

70. MacArthur, R. H., and J. W. MacArthur, "On Bird Species Diversity," Ecology, Vol 42 (1961), pp 594-598.

71. Morrison, D. F., Multivariate Statistical Methods (McGraw Hill, New York, NY, 1976).

72. Neuhaus, J. O., and C. Wrigley, "The Quartimax Method: An Analytical Approach to Orthogonal Simple Structure," Brit. J. Stat. Psych., Vol 7 (1954), pp 81-91.

73. Noon, B. R., D. K. Dawson, D. B. Inkley, C. S. Robbins, and S. H. Anderson, "Consistency in Habitat Preference of Forest Bird Species," Trans. N. Amer. Wildl. Nat. Res. Conf., Vol 45 (1980), pp 226-244.

74. Noy-Meir, I., "Data Transformations in Ecological Ordination. I. Some Advantages of Non-Centering," J. Ecol., Vol 61 (1973), pp 329-341.

75. Orloci, L., Multivariate Analysis in Vegetation Research, 2nd ed. (Junk Publishers, The Hague, Netherlands, 1978).

76. Pielou, E. C., Mathematical Ecology (John Wiley & Sons, New York, NY, 1977).

77. Pierce, V., and T. C. Grubb, Jr., "Laboratory Studies of Foraging in Four Bird Species of Deciduous Woodland," Auk, Vol 98 (1981), pp 307-320.

78. Pimentel, R. A., Morphometrics: The Multivariate Analysis of Biological Data (Kendall/Hunt, Dubuque, IA, 1979).

79. Rao, C. R., Linear Statistical Inference (John Wiley & Sons, 1973).

80. Recher, H. F., "Bird Species Diversity and Habitat Diversity in Australia and North America," Amer. Natur., Vol 103 (1969), pp 75-80.

81. Rhoades, D. F., and R. G. Cates, "Toward a General Theory of Plant Antiherbivore Chemistry," Rec. Adv. Phytochem., Vol 10 (1976), pp 168-213.

82. Rohlf, F. J., "Classification of Aedes by Numerical Taxonomic Methods (Diptera: Culicidae)," Ann. Entomol. Soc. Amer., Vol 56 (1963), pp 798-804.

83. Rohlf, F. J., "Methods of Comparing Classifications," Ann. Rev. Ecol. Syst., Vol 5 (1974), pp 101-113.

84. Romesburg, H. C., Cluster Analysis for Researchers (Lifetime Learning Publications, Belmont, CA, 1984).

85. Rosenthal, G. A., and D. H. Janzen, Eds., Herbivores. Their Interaction With Secondary Plant Metabolites (Academic Press, New York, NY, 1979).

86. Roth, R. R., "Spatial Heterogeneity and Bird Species Diversity," Ecology, Vol 57 (1976), pp 773-782.

87. SAS User's Guide: Statistics, 1982 ed. (SAS Institute Inc., Cary, NC, 1982).

88. Saunders, D. R., An Analytic Method for Rotation to Orthogonal Simple Structure, Research Bulletin 53-10 (Educational Testing Service, Princeton, NJ, 1953).

89. Searle, S. R., Matrix Algebra Useful for Statistics (John Wiley & Sons, New York, NY, 1982).

90. Seber, G. A. F., Multivariate Observations (John Wiley & Sons, New York, NY, 1984).

91. Shugart, H. H., and B. C. Patten, "Niche Quantification and the Concept of Niche Pattern," in B. C. Patten, Ed., Systems Analysis and Simulation in Ecology, Vol II (Academic Press, New York, NY, 1972), pp 283-327.

92. Simpson, G. G., "Notes on the Measurement of Faunal Resemblance," Amer. J. Sci., Vol 258 (1960), pp 300-311.

93. Sneath, P. H. A., "Evaluation of Clustering Methods," in A. J. Cole, Ed., Numerical Taxonomy. Proceedings of the Colloquium in Numerical Taxonomy, held at the University of St. Andrews, September 1968 (Academic Press, London, 1969).

94. Sneath, P. H. A., and R. R. Sokal, Numerical Taxonomy: The Principles and Practice of Numerical Classification (W. H. Freeman and Co., San Francisco, CA, 1973).

95. Sokal, R. R., and C. D. Michener, "A Statistical Method for Evaluating Systematic Relationships," Univ. Kansas Sci. Bull., Vol 38 (1958), pp 1409-1438.

96. Sokal, R. R., and F. J. Rohlf, Biometry (W. H. Freeman and Co., San Francisco, CA, 1969).

97. Southwood, T. R. E., "The Insect/Plant Relationship--An Evolutionary Perspective," Symp. R. Entomol. Soc. Lond., Vol 6 (1973), pp 3-30.

98. SPSS^{x} User's Guide, 1985 Ed. (SPSS Inc., Chicago, IL, 1985).

99. Sternberg, L., "Growth Forms of Larrea tridentata," Madroño, Vol 23 (1976), pp 408-417.

100. Swan, J. M. A., "An Examination of Some Ordinating Problems by Use of Simulated Vegetational Data," Ecology, Vol 51 (1970), pp 89-101.

101. Tatsuoka, M. M., Discriminant Analysis. The Study of Group Differences (Institute for Personality and Ability Testing, Champaign, IL, 1970).

102. Tatsuoka, M. M., Multivariate Analysis: Techniques for Educational and Psychological Research (John Wiley & Sons, New York, NY, 1971).

103. Thurstone, L. L., The Vectors of Mind (Univ. Chicago Press, Chicago, IL, 1935).

104. Thurstone, L. L., "A New Rotational Method in Factor Analysis," Psychol., Vol 3 (1938), pp 199-218.

105. Thurstone, L. L., Multiple Factor Analysis (Univ. Chicago Press, Chicago, IL, 1947).

106. Tomoff, C. S., "Avian Species Diversity in Desert Scrub," Ecology, Vol 55 (1974), pp 396-403.

107. Travis, J., "Seasonal Foraging in a Downy Woodpecker Population," Condor, Vol 79 (1977), pp 371-375.

108. Van Ness, J. W., and C. Simpson, "On the Effects of Dimension in Discriminant Analysis," Technometrics, Vol 18 (1976), pp 175-187.

109. Vasek, F. C., "Creosote Bush: Long-Lived Clones in the Mohave Desert," Amer. J. Bot., Vol 67 (1980), pp 246-255.

110. Wahl, P. W., and R. A. Kronmal, "Discriminant Functions When Covariances Are Unequal and Sample Sizes Are Moderate," Biometrics, Vol 33 (1977), pp 479-484.

111. Whitmore, R. C., "Habitat Ordination of Passerine Birds of the Virgin River Valley, Southwestern Utah," Wilson Bull., Vol 87 (1975), pp 65-74.

112. Whitmore, R. C., "Habitat Partitioning in a Community of Passerine Birds," Wilson Bull., Vol 89 (1977), pp 253-265.

113. Whittaker, R. H., Ed., Ordination of Plant Communities (Junk Publishers, The Hague, Netherlands, 1978).

114. Wiens, J. A., and J. T. Rotenberry, "Habitat Associations and Community Structure of Birds in Shrubsteppe Environments," Ecol. Monogr., Vol 51 (1981), pp 21-41.

115. Williams, B. K., "Discriminant Analysis in Wildlife Research: Theory and Applications," in D. E. Capen, Ed., The Use of Multivariate Statistics in Studies of Wildlife Habitat, U. S. For. Serv., Gen. Tech. Rep. RM-87 (U.S. Forest Service, Fort Collins, CO, 1981), pp 59-71.

116. Williams, B. K., "Some Observations on the Use of Discriminant Analysis in Ecology," Ecology, Vol 64 (1983), pp 1283-1291.

117. Williams, W. T., and J. M. Lambert, "Multivariate Methods in Plant Ecology: V. Similarity Analyses and Information Analysis," J. Ecol., Vol 54 (1966), pp 427-445.

118. Willson, M. F., "Foraging Behavior of Some Winter Birds of Deciduous Woods," Condor, Vol 72 (1970), pp 169-174.

119. Willson, M. F., "Avian Community Organization and Habitat Structure," Ecology, Vol 55 (1974), pp 1017-1029.

120. Wishart, D., CLUSTAN User Manual (Program Library Unit, Edinburgh Univ., Edinburgh, Scotland, 1981).

APPENDIX A:

ECOLOGICAL CHARACTERISTICS OF THE MOJAVE DESERT STUDY SITES

Code	Habitat	Environmental Impact	Relative Disturbance (% Vehicle Imprints)	Size (ha)	Number of Transects	Shrub Volume (m^3/ha)	Creosote Bush Cover (m^2/ha)	Burrobush Cover (m^2/ha)	Shrub Cover (m^2/ha) of Five Height Categories: < 0.5 m	0.5 to < 1 m	1 to < 1.5 m	> 1.5 m	Substrate Particle Sizes (%): < 3 mm	3 mm to < 1 cm	1 to < 8 cm	Ground Cover (%): GRASS	FORBS
[illegible]	Valley	Severe	86	96	6	198	187	25	62	34	71	19	82	13	4.4	18	5.5
M	Alluvial	Moderate	39	96	6	525	343	154	148	171	144	114	47	24	24	20	4.5
[illegible]	High Desert Plateau	Light	8	64	4	1194	837	147	119	238	444	283	6.5	25	64	7.4	20
A[illegible]	Alluvial	None	0	48	3	1813	1296	455	359	398	866	243	3.2	19	76	15	28
V[illegible]	Valley	None	[illegible]	64	4	2202	1317	637	273	476	536	603	29	63	8.6	27	12

All Data Collected [illegible] April 198[illegible].

APPENDIX B:

MOJAVE HABITAT VARIABLES

Number	Code	Habitat Variable	
1	CCB	cover of Larrea tridentata (creosote bush)	(m^2/ha)
2	CBW	cover of Ambrosia dumosa (burroweed)	(m^2/ha)
3	CCHE	cover of Hymenoclea salsola	(m^2/ha)
4	CSH	cover of Grayia spinosa + Lycium spp.	(m^2/ha)
5	CTB	cover of Thamnosma montana	(m^2/ha)
6	CMT	cover of Ephedra spp.	(m^2/ha)
7	CCAS	cover of Cassia armata	(m^2/ha)
8	CMD	cover of Psorothamnus arborescens	(m^2/ha)
9	CA	cover of Acamptopappus sphaerocephalus	(m^2/ha)
10	CCD	cover of Salazaria mexicana + Eriogonum fasciculatum	(m^2/ha)
11	C123	cover of Amphipappus fremontii + Haplopappus cooperi + Encelia frutescens	(m^2/ha)
12	CREST	cover of CCHE to C123 combined (subdominant shrubs)	(m^2/ha)
13	HCB	mean height of Larrea tridentata	(cm)
14-24	HBW-HREST	mean height of HBW-HREST	(cm)
25	CL1	cover of shrubs < 0.5 m in height	(m^2/ha)
26	CL2	cover of shrubs $\geq$ 0.5 m to < 1 m in height	(m^2/ha)
27	CL3	cover of shrubs $\geq$ 1 m to < 1.5 m in height	(m^2/ha)
28	CL4	cover of shrubs $\geq$ 1.5 m in height	(m^2/ha)
29	GRASS	grass cover	(%)
30	FORB	forb cover	(%)
31	LIT	litter cover	(%)
32	SAND	sand (particle sizes < 3 mm)	(%)
33	CSAND	coarse sand (particle sizes $\geq$ 3 mm to < 1 cm)	(%)
34	GRAV	gravel (particle sizes $\geq$ 1 to < 8 cm)	(%)
35	ROCK	rock (particle sizes $\geq$8 cm)	(%)
36	CVDCB	coefficient of variation, diameter of creosote bush	
37	CVDBW	coefficient of variation, diameter of burroweed	
38	CVDR	coefficient of variation, diameter of subdominant shrubs	
39	CVHCB	coefficient of variation, height of creosote bush	
40	CVHBW	coefficient of variation, height of burroweed	
41	CVHR	coefficient of variation, height of subdominant shrubs	
42	CVNCB	coefficient of variation, density of creosote bush	
43	CVNBW	coefficient of variation, density of burroweed	
44	CVNR	coefficient of variation, density of subdominant shrubs	
45	L2	density of shrubs 0.2 m in height	$(shrubs/1600\ m^2)$
46	L4	density of shrubs 0.3 - 0.4 m in height	$(shrubs/1600\ m^2)$
47	L6	density of shrubs 0.5 - 0.6 m in height	$(shrubs/1600\ m^2)$
48	L8	density of shrubs 0.7 - 0.8 m in height	$(shrubs/1600\ m^2)$
49	L10	density of shrubs 0.9 - 1.0 m in height	$(shrubs/1600\ m^2)$
50	L12	density of shrubs 1.1 - 1.2 m in height	$(shrubs/1600\ m^2)$
51	L14	density of shrubs 1.3 - 1.4 m in height	$(shrubs/1600\ m^2)$
52	L16	density of shrubs 1.5 - 1.6 m in height	$(shrubs/1600\ m^2)$
53	L18	density of shrubs 1.7 - 1.8 m in height	$(shrubs/1600\ m^2)$
54	L20	density of shrubs 1.9 - 2.0 m in height	$(shrubs/1600\ m^2)$
55	L22	density of shrubs 2.1 - 2.2 m in height	$(shrubs/1600\ m^2)$
56	L27	density of shrubs > 2.2 m in height	$(shrubs/1600\ m^2)$
57	SHCOV	total shrub cover	(m^2/ha)
58	NSP	number of shrub species	(N)
59	L4J*	L4 + L6	$(shrubs/1600\ m^2)$
60	L12J*	L12 + L14	$(shrubs/1600\ m^2)$

*Variables created after initial data analysis because of very high correlations.

APPENDIX C:

VARIABLE TRANSFORMATIONS USED FOR MOJAVE DESERT HABITAT VARIABLES

Variable (X)	Transformation of Variable X
Habitat variables:	
CCB to crest and L2 to SHCOV (see Appendix B)	$\log_e(X + 1)$
HCB to HREST (see Appendix B)	$\log_e(100X + 1)$
Shrub diameter and height	$\log_e(100X)$
Shrub density	$X^{1/2}$
Ground cover and substrate size	arcsine p_x*
Horizontal heterogeneity CVDCB to CVNR (see Appendix B)	$\log_e(100X)$
Population estimates for mammals	$\log_e(X + 1)$
Population estimates for birds	$(X + 0.5)^{1/2}$
Biomass	$\log_e(X)$

*Variable X is expressed as a proportion (decimal). Since the arcsine transformation is unavailable with many statistical packages, an equivalent form was used:

$$\arctan[p_x/(1 - p_x^2)^{1/2}]$$

APPENDIX D:

ECOLOGICAL CHARACTERISTICS OF THE ILLINOIS STUDY SITES

Code	Habitat	Age (Years)	Size (ha)	Number of Quadrats	Ground Cover (%)	Shrub/Seedling Density (Stems/0.1 ha)	Canopy Cover (%)	Mean Vegetation Height (m)	Mean Max. Canopy Height (m)	Max. Canopy Height (m)
A*	Abandoned Pasture	10-12[1]	10	40	96	113	4.9	1.5	3.8	6.1
B	Herbaceous	16[1]	10	36	64	338	6.1	1.1	6.7	15.2
C	Young Locust[3]	18 (10-12 most woody vegetation)	10	36	64	139	48	5.0	8.6	15.0
D	Decadent Locust[3]	21[1]	10	36	61	421	50	7.6	13.2	17.0
E	Late Old Field[3]	28[1]	10	36	70	258	58	9.7	19.4	25.0
F	Retarded Old Field[3]	57[1,2]	7.8	28	70	622 ■	27 ■	2.7 ■	11.2 ■	22.9
G	Bottomland Forest[3]	50[1,2]	9.7	35	58	162	88	19.4	32.9	41.1
H	Bottomland Forest[3]	80[2]	10	36	51	85	89	24.0	32.3	41.1
I	Upland Forest (Steep Ravines— Some Riparian)	Complex (40-100+)[3]	10.3	37	47	264	92	23.9	33.5	51.8
J	Upland Forest	100+[3]	10	36	50	233	87	26.3	32.7	44.2

Number of Quadrats = 356

A* Phillips Tract abandoned pasture, not including sapling edge habitat.

Abandoned strip-mines.

■ Mosaic of dense shrubs and open standing water, with scattered large trees.

1 Determined from tree corings

2 Determined from strip-mining records

3 Estimated

APPENDIX E:

ILLINOIS HABITAT VARIABLES

Variable	Type of Variable		
L1	S	Vegetation hits	0 - < 1 m
L2	S	Vegetation hits	1 - < 2 m
L3	S	Vegetation hits	2 - < 3 m
L4	S	Vegetation hits	3 - < 4 m
L5	S	Vegetation hits	4 - < 5 m
L6	S	Vegetation hits	5 - < 6.5 m
L7	S	Vegetation hits	6.5 - < 7.5 m
L8	S	Vegetation hits	7.5 - < 8.5 m
L9	S	Vegetation hits	8.5 - < 10 m
L10	S	Vegetation hits	10 - < 15 m
L11	S	Vegetation hits	≥ 15 m
VEGVOL	S	Total vegetation hits	
VHW	S	Vertical heterogeneity within quadrat	
HZHW	S	Horizontal heterogeneity within quadrat	
GC	S,T	Ground cover (%)	
CC	S,T	Canopy cover (%)	
AQUA	S	Aquatic hits (%)	
TOP	S	Level ground (%)	
SHRD	T	Shrub density (stems/0.1 ha)	
SHRH	T	Shrub heterogeneity	
SHRC	T	Shrub cover (m^2/0.1 ha)	
NSPSHR	T	Number of shrub species in the 16 individuals measured	
CANAV	T	Average height of dominant vegetation (m)	
CANMX	T	Highest vertical projection of vegetation (m)	
NSPTRE	T	Number of tree species found in the 26-m quadrat	
SAP	T	Density of trees 3 to < 8 cm dbh	(number/ha)
A	T	Density of trees 8 to < 15 cm dbh	(number/ha)
B	T	Density of trees 15 to < 22 cm dbh	(number/ha)
C	T	Density of trees 22 to < 30 cm dbh	(number/ha)
D	T	Density of trees 30 to < 38 cm dbh	(number/ha)
E	T	Density of trees 38 to < 53 cm dbh	(number/ha)
F	T	Density of trees 53 to < 68 cm dbh	(number/ha)
G	T	Density of trees 68 to < 84 cm dbh	(number/ha)
H	T	Density of trees 84 cm to < 1 m dbh	(number/ha)
I	T	Density of trees ≥ 1 m dbh	(number/ha)
TREMAS	T	Total basal area of trees	(m^2/0.053 ha)
DUP	G	Basal area of upland tree species	(m^2/0.053 ha)
DBOT	G	Basal area of bottomland tree species	(m^2/0.053 ha)
DRICH	G	Basal area of rich/moist soil tree species	(m^2/0.053 ha)
DEAR	G	Basal area of early succession tree species	(m^2/0.053 ha)
DDEA	G	Basal area of snags	
DBL	G	Basal area of black locust	
DC	G	Basal area of cottonwood	
DELM	G	Basal area of American and slippery elms	

*S=structural; T=typical; G=tree guilds.

APPENDIX F:

ILLINOIS TREE SPECIES, SCIENTIFIC NAMES, AND GUILD COMPOSITIONS

Cosmopolitan	Black locust	Robinia pseudoacacia
	Cottonwood	Populus deltoides
Elms	American elm	Ulmus americana
	Siberian elm	Ulmus pumila
	Slippery elm	Ulmus rubra
Upland Species	White ash	Fraxinus americana
	Basswood	Tilia americana
	Flowering dogwood	Cornus florida
	Shagbark hickory	Carya ovata
	Sweet pignut hickory	
	Hop hornbeam	Ostrya virginiana
	Sugar maple	Acer saccharum
	Red mulberry	Morus rubra
	White mulberry	Morus alba
	Black oak	Quercus velutina
	Chinquapin oak	Quercus muehlenbergii
	White oak	Quercus alba
	Blackhaw	Viburnum prunifolium
Bottomland Species	European black alder	Alnus glutinosa
	Green ash	Fraxinus pennsylvanica
	Bald cypress	Taxodium distichum
	Burning bush	Euonymus atropurpureus
	Gray dogwood	Cornus racemosa
	Rough-leafed dogwood	Cornus drummondii
	Sweetgum	Liquidambar styraciflua
	Box elder	Acer negundo
	Silver maple	Acer saccharinum
	Bur oak	Quercus macrocarpa
	Swamp white oak	Quercus bicolor
	Sycamore	Plantanus occidentalis
	Willow	Salix spp.
Moist/Rich Soil Species	Ohio buckeye	Aesculus glabra
	Catalpa	Catalpa speciosa
	Kentucky coffee tree	Gymnocladus dioicus
	Hackberry	Celtis occidentalis
	Bitternut hickory	Carya cordiformis
	Ironwood	Carpinus caroliniana
	Nannyberry	Viburnum lentago
	Red oak	Quercus rubra
	Shingle oak	Quercus imbricaria
	Paw paw	Asimina triloba
	Northern prickly ash	Zanthoxylum americanum
	Redbud	Cercis canadensis
	Black walnut	Juglans nigra
	Butternut	Juglans cinerea
Early Succesion Species	Eastern red cedar	Juniperus virginiana
	Black cherry	Prunus serotina
	Crabapple	Malus spp.
	Hawthorn	Crataegus spp.
	Honey locust	Gleditsia triacanthos
	Osage orange	Maclura pomifera
	Peach	Prunus persica
	Plum	Prunus americana
	Sassafras	Sassafras albidum
	Smooth sumac	Rhus glabra

APPENDIX G:

CHARACTERIZATION OF DISPERSION AND COVARIANCE MATRICES

The dispersion matrix consists of elements of the form:

$$\sum_{i=1}^{N} (X_{\alpha i} - \bar{X}_{\alpha})(X_{\beta i} - \bar{X}_{\beta}) \qquad \text{[Eq G1]}$$

$$\alpha = 1,2,\ldots\ldots p$$

$$\beta = 1,2,\ldots\ldots p$$

This matrix, which is also called the sums of squares and cross products matrix (SSCP), gives information about the scatter of p variates around their means with a sample size N.

However, the magnitude of the elements is a function of sample size. It is obvious that as N increases, the above term increases. Dividing by sample size N gives each element an "average dispersion" and the matrix, called a variance-covariance (or more simply covariance) matrix, is independent of sample size.

Since $\alpha = \beta$ along the diagonal of the matrix, diagonal elements will be:

$$1/N \sum_{i=1}^{N} (X_{\alpha i} - \bar{X}_{\alpha})^2.$$

Since this term is recognized as the variance, a covariance matrix consists of p variance components along its diagonal, and $(p^2 - p)/2$ covariances as off-diagonal elements.

USA-CERL DISTRIBUTION

Chief of Engineers 20314
ATTN: Tech Monitor
ATTN: DAEN-ZCF-U
ATTN: CECW-ZR (3)
ATTN: CECW-R (2)
ATTN: CEEC-EI (2)
ATTN: CEEC-ET (2)
ATTN: CEEC-M (2)

US Army Engineering District (13)
ATTN: Military Planning Section*
New York 10278
Baltimore 21203
Norfolk 23510
Savannah 31402
Omaha 68102
Fort Worth 76102
Los Angeles 90053
Sacramento 95814
Huntsville 35807
ATTN: HNDAD-O
Mobile 36628
ATTN: SAMEN-MP
Rock Island 55101
ATTN: Regulatory Functions
Kansas City 64106
ATTN: Inst. Plan & Support Sec.
Alaska 99506
ATTN: NPAEN-PM-AF

US Military Academy 10996
ATTN: Dept of Mechanics
ATTN: Library
ATTN: Envr Ofc

The Army Library (ANRAL-R) 20310
ATTN: Army Studies Section

AMC 22333
ATTN: AMXMC-D (2)

US Army HQ FORSCOM 30330
ATTN: AFEN-EQ (4)
ATTN: AFOP-TS

Ft. Belvoir, VA 22060
ATTN: ATSEN-DT-LD (2)
ATTN: Archives Section/Bldg 270
ATTN: ATZA-DEH-EN
ATTN: STRBE-U

Director, USA-WES 39180
ATTN: CEWES-ER
ATTN: CEWES-EN

National Guard 20310
ATTN: NGB-ARO-AM

Env. Protection Agency (EPA) 20460

Fort Irwin 92310
ATTN: AFZJ-EHE-S
ATTN: AMXLA-SA

Wright Patterson AFB 45433
ATTN: HQAFLC-DEVX

Edwards AFB 93523
ATTN: AFFTC/DEV (2)

Defense Technical Info. Ctr. 22314
ATTN: DDA (2)

62
+5

END
Feb.
1988
DTIC